AD-A190 199

1a. REPORT SECURITY CLASSIFICATION
Unclassified

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for Public Release; Distribution Unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
Annual Report No. 37

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
Columbia University
Radiation Laboratory

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
Department of the Army

6c. ADDRESS *(City, State, and ZIP Code)*
Columbia Radiation Laboratory
Columbia University
New York, NY 10027

7b. ADDRESS *(City, State, and ZIP Code)*
U.S. Army Research Office
Research Triangle Park, NC 27709

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
Department of the Army

8b. OFFICE SYMBOL *(If applicable)*

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
DAAG29-85-K-0049

8c. ADDRESS *(City, State, and ZIP Code)*
U.S. Army Research Office
Research Triangle Park, NC 27709

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE *(Include Security Classification)*
RESEARCH INVESTIGATION DIRECTED TOWARD EXTENDING THE USEFUL RANGE OF THE ELECTROMAGNETIC SPECTRUM

12. PERSONAL AUTHOR(S)
George W. Flynn, Richard M. Osgood, Jr.

13a. TYPE OF REPORT
Annual

13b. TIME COVERED
FROM 10/1/86 TO 9/30/87

14. DATE OF REPORT *(Year, Month, Day)*
December 31, 1987

15. PAGE COUNT

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
photon-number-squeezed light
theory of light detection,
excess noise factor

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The usual semiclassical theory of light detection has been demonstrated to be valid only in the absence of feedback from detector to source. A revised theory has been developed that is valid even in the presence of such feedback. Research has begun on the possibility of using various heterostructures (including quantum-well devices) that might serve as solid-state versions of the space-charge-limited Franck-Hertz experiment (this experiment produced the first source of cw photon-number-squeezed light, reported in 1985.) →

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
George W. Flynn

22b. TELEPHONE *(Include Area Code)*
(212) 280-3265

22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR
83 APR edition may be used until exhausted.
All other editions are obsolete.

Block 18 continued--Subject Terms (Key Words)

feedback
sub-Poisson light   cosmic-ray noise
superlattice avalanche photodiodes   Cerenkov radiation
$Cu+Cl_2$ Laser Induced Desorption
Time-Of-Flight   mass spectrometer
ion ejection $CuCl_x$
Excimer Si
Ge   Quadropole mass spectrometer
ions   photoionic emission
threshold   low temperature processing
low energy ion beam   silicon oxidation
sputtering of silicon dioxide   germanium nitride
gallium arsenide MESFET   gate leakage current
Collisions   Hot Atoms
Carbon Dioxide   Hydrogen
Deuteruim   Diode Laser
Infrared Absorption   Excimer Laser
Pyruvic Acid   Energy Transfer
Photodissociation   Chlorine atoms
Surface   Picosecond
Femtosecond   Second-Harmonic
Interface   Isomerization
Binaphthyl   Water
Liquid   Erythrosin
Adsorption   Isotherm
Structure   Capacitance
Capacitance Measurement   Negative Capacitance
Schottky Capacitance   Schottky Barrier
Interface States   Metal-semiconductor Interface
$NiSi_2-Si$ Impact Ionization
Optical Coherent Transients   Photon Echoes
Sodium Vapor   Spectroscopy
Ultrafast Modulation Spectroscopy   Non-linear Optics
Four-Wave Mixing   Time-Delayed Four-Wave Mixing
Photon Echo Modulation Spectroscopy   Etching
In Situ Processing   Photoelectron Emission

PUBLICATIONS AND REPORTS

K. B. Eisenthal, K. Bhattacharyya, and E. V. Sitzmann, "Study of Chemical Reactions by Surface Second Harmonic Generation: p-Nitrophenol at the Air-Water Interface," **J. Chem. Phys.** 87(1987):1442.

T. Kreutz, J. O'Neill, and G. Flynn, "IR Diode Laser Study of Vibrational Energy Distribution in CO_2 Produced by UV Excimer Laser Photofragmentation of Pyruvic Acid," **J. Chem. Phys.** 87(1987):4598-4605.

A. Hewitt, J. Hershberger, G. Flynn, and R. Weston, Jr., "Rotationally Resolved Isotope Effect in the Hot Atom Collisional Excitation of CO_2 (00^01) by Time-Dependent Diode Laser Spectroscopy," **J. Chem. Phys.** 87(1987):1894.

T. Kreutz, J. O'Neill, and G. Flynn, "Diode Laser Absorption Probe of V-V Energy Transfer in CO_2," **J. Phys. Chem.** 91(1987):5540-5543.

S. S. Todorov, C. F. Yu, and E. R. Fossum, "Direct Formation of Dielectric Thin Films on Silicon by Low Energy Ion Beam Bombardment," **Vacuum** 36(1986):929-932.

C. F. Yu, S. S. Todorov, and E. R. Fossum, "Characterization of Ultra-Thin SiO_xO_2 Films Formed by Direct Low Energy Ion Beam Oxidation," **J. Vac. Sci. Technol.** A5(4)(1987):1569-1571.

D. V. Rossi, E. R. Fossum, G. D. Pettit, P. D. Kirchner, and J. M. Woodall, "Reduced Reverse Bias Current in Al-GaAs and $In_{0.75}Ga_{0.25}$As-GaAs Junctions Containing an Interfacial Arsenic Layer," **J. Vac. Sci.Technol.** B5(1987):982-984.

S. S. Todorov and E. R. Fossum, "Growth Mechanism of Thin Oxide Films Under Low Energy Oxygen Ion Bombardment," to be published in **J. Vac. Sci. Technol.**

S. S. Todorov and E. R. Fossum, "Sputtering of Silicon Dioxide Near Threshold," to be published in **Appl. Phys. Lett.**

S. S. Todorov and E. R. Fossum, "Oxidation of Silicon by a Low Energy Ion Beam: Model and Experiment," to be published in **Appl. Phys. Lett.**

S. S. Todorov, "Interactions of Low Energy Oxygen Ions with Silicon Surfaces," Ph. D. diss., Columbia University, 1987.

S. R. Hartmann, "Using Incoherent Light to Generate Coherent Excitations," Proceedings of the International Laser Science Conference ILS-II, Seattle, Washington, 1986, to be published in **Am. Inst. Phys. Conf. Proc.**

D. DeBeer, S. R. Hartmann and R. Friedberg, "Comment on Diffraction Free Beams," to be published in **Phys. Rev. Lett.**

R. Friedberg and S. R. Hartmann,"A Diagramatic Technique for Calculating Radiation of Coherently or Incoherently Excited Two-level Atoms," to be published in **J. Phys. B.**

H. S. Cho and P. R. Prucnal, "New Formalism of the Kronig-Penney Model with Application to Superlattices," **Phys. Rev. B** 36(1987):3237-3242.

W. Holber, J. O. Chu, D. Gaines, A. Nahata, and R. M. Osgood, "Laser Assisted Plasma Etching," ECS Proceedings, 1986.

C. F. Yu, M. T. Schmidt, D. V. Podlesnik, and R. M. Osgood, "Optically-Induced, Room-Temperature Oxidation of Gallium Arsenide," **Mat. Res. Soc. Symp. Proc.** 75(1987):251-255.

P. D. Brewer and R. M. Osgood,"Large Area Laser-Assisted Etching of Electronic Materials," SPIE 611(1986):62.

R. W. Ade, E. E. Harstead, T. Cacouris, E. R. Fossum, P. R. Prucnal, and R. M. Osgood, "Direct Connection of Optical Fibers to Integrated Circuits," IEPS Proceedings, San Diego, 1986.

C. F. Yu, M. T. Schmidt, D. V. Podlesnik, and R. M. Osgood, Jr., "Wavelength Dependence of Optically Induced Oxidation of GaAs (100)," **J. Vac. Sci. Technol.** B5(1987):1087-1091.

W. Holber, D. Gaines, C. F. Yu, R. M. Osgood, "Laser Desorption of Polymer in a Plasma Reactor," to be published in **Appl. Phys. Lett.**

D. V. Podlesnik, "Light-Guided Etching for III-V Semiconductor Device Fabrication," Proceedings of the European Solid State Device Research Conference, Bologna, Italy, (1987): 462-470.

C. F. Yu, M. T. Schmidt, D. V. Podlesnik, E. S. Yang, and R. M. Osgood, Jr., "Ultraviolet-Light-Enhanced Reaction of Oxygen with Gallium Arsenide Surfaces," to be published in **J. Vac. Sci. Tech.**

X. Wu, H. L. Evans, E. S. Yang and P. S. Ho, "An improved Differential Voltage Technique for Capacitance Measurement," to be published in **Solid State Electronics.**

M. T. Schmidt, D. V. Podlesnik, H. L. Evans, C. F. Yu, E. S. Yang, and R. M. Osgood, Jr., "The Effect of a Thin UV-Grown Oxide on Metal-GaAs Contacts," to be published in **J. Vac. Sci. Technol.**

M. C. Teich, Review of R. P. Feynman's book entitled, "Surely You're Joking, Mr. Feynman! Adventures of a Curious Character," **Phys. Today** 39, #9(1986):61.

J. H. Shapiro, P. Kumar, G. Saplakoglu, M. C. Teich, and B. E. A. Saleh, "Theory of Light Detection in the Presence of Feedback," J. Opt. Soc. Am. B 3(1986):66.

M. C. Teich, K. Matsuo, and B. E. A. Saleh, "Excess Noise Factor and Gain Distributions for Superlattice Avalanche Photodiodes," J. Opt. Soc. Am. A 3(1986):39.

F. Capasso and M. C. Teich, "Conversion of Poisson Photons into Sub-Poisson Photons by the Action of Electron Feedback," Phys. Rev. Lett. 57(1986):1417-1420.

M. C. Teich, K. Matsuo, and B. E. A. Saleh, "Excess Noise Factors for Conventional and Superlattice Avalanche Photodiodes and Photomultiplier Tubes," IEEE J. Quant. Electron QE-22(1986):1184-1193.

M. C. Teich, F. Capasso, and B. E. A. Saleh, "Photon-Number-Squeezed Recombination Radiation in Semiconductors," J. Opt. Soc. Am. B 4(1987):1663-1666.

J. H. Shapiro, G. Saplakoglu, S. T. Ho, P. Kumar, B. E. A. Saleh, and M. C. Teich, "Theory of Light Detection in the Presence of Feedback," J. Opt. Soc. Am. B 4(1987):1604-1620.

M. C. Teich and B. E. A. Saleh, "Approximate Photocounting Statistics of Shot-Noise Light with Arbitrary Spectrum," J. Mod. Opt. 34(1987):1169-1178.

B. E. A. Saleh and M. C. Teich, "Can the Channel Capacity of a Lightwave Communication System be Increased by the Use of Photon-Number-Squeezed Light," Phys. Rev. Lett. 58(1987):2656-2659.

M. C. Teich, R. A. Campos, and B. E. A. Saleh, "Statistical Properties of Clustered Cosmic-Ray Events at Ground Level Determined from Photomultiplier-Tube Background Registrations," Phys. Rev. D(1987):2649-2665.
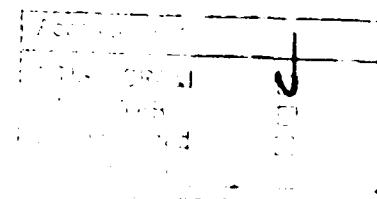
A-1

3

Study of chemical reactions by surface second harmonic generation: p-Nitrophenol at the air–water interface

Kankan Bhattacharyya, E. V. Sitzmann, and K. B. Eisenthal

Department of Chemistry, Columbia University, New York, New York 10027

In recent years, optical second harmonic generation (SHG) has been shown to provide valuable new information on a number of interfacial properties.[1] Since SHG is electric dipole forbidden in the bulk, this technique can be used to probe selectively the interfacial layer between two centrosymmetric media. Thus a wide variety of surfaces have been studied by this technique; a substantial amount of knowledge has been gathered about concentration, orientation, and spectroscopy of the species at the interfaces.[2-13] In our earlier work, we focused our attention on the relative[12] and absolute[13] orientation of molecules at the surface of liquid solutions. We now address the question of how the asymmetry in the forces experienced by the molecules in the interfacial region affects simple chemical equilibria. Specifically, using this technique we have investigated the acid–base equilibrium between p-nitrophenol (HA) and its anion (A⁻).

$$\begin{array}{c}\text{OH} \\ \bigcirc \\ \text{NO}_2\end{array} + H_2O \rightleftharpoons \begin{array}{c}\text{O}^{(-)} \\ \bigcirc \\ \text{NO}_2\end{array} + H_3O^{(+)} \qquad (1)$$
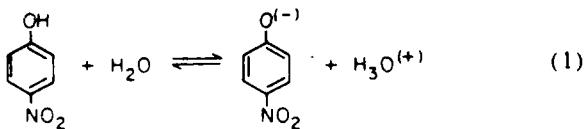
As we will show, not only do the concentrations of the various species in the interfacial region differ from their bulk values, but also the relative concentrations do not correspond to the bulk equilibrium constant. The "equilibrium constant" in the surface region is vastly different from that in the bulk. The key factor is most likely the reduced polarity at the surface due to the low density of molecules on the vapor side of the surface. The effect of this is to decrease the stability of charged species, e.g., A⁻, to such an extent that the equilibrium in the interfacial region overwhelmingly favors the neutral undissociated HA molecules. It is to be noted that although the concentrations are dependent on their position, varying in a continuous, though very sharp way, from the bulk vapor to the bulk liquid, we can still define an average concentration for each of the species in this inhomogeneous region. We then seek to compare the relative average concentrations of the species in the interfacial region, e.g., A⁻/HA, with the relative concentrations given by the bulk equilibrium constant. In order to treat the position-dependent concentrations (activities) in the interfacial region, we would use the chemical potential, since its value for a given species is the same at all positions in the system, i.e., bulk liquid, vapor, and the inhomogenous interfacial region. Discussions of this approach will be postponed to a later publication.

The experimental setup consisting of a cavity-dumped, synchronously pumped dye laser, and single photon counting electronics is described elsewhere.[11] The component of the surface second order nonlinear susceptibility $\overrightarrow{\chi_s^{(2)}}$ per-

pendicular to the plane of incidence was measured by setting the analyzer to collect only the s-polarized SH light. Such an arrangement is selected because it contains the contribution from the surface region alone and not the bulk region.[2]

The observed s-polarized SH field amplitude ($E_s^{2\omega}$) as a function of bulk pH of the solution is shown in Fig. 1. We found in earlier work[13] that up to a monolayer the nonlinear susceptibility $\overrightarrow{\chi_s^{(2)}}$ can be written as

$$\overrightarrow{\chi_s^{(2)}} = N_s \langle \tilde{\alpha}^{(2)} \rangle, \qquad (2)$$

where N_s is the number of molecules per unit area of the surface and $\tilde{\alpha}^{(2)}$ the molecular polarizability. From this relation it follows that the second harmonic field $E_s^{2\omega}$ is linear in N_s. As the bulk pH is varied, the relative amount of HA and A⁻, as given by the bulk equilibrium constant for reaction (1) changes. This results in a change in N_s as the latter depends on bulk concentrations. At low pH, the bulk concentration of HA has its largest value and this leads to a high surface coverage of neutral HA molecules. In accordance with this an intense signal (about 80 times that of water) was observed. In extremely alkaline media (pH ≥ 10), the bulk HA molecules are almost totally converted into the anion and we found that the SH signal decreased to the value obtained from pure water. If there was a significant anion concentration at the surface we would readily observe it since its SH signal should be more than 10 times that of water. The possibility of undissociated sodium phenolate formed by neutralization with NaOH being present at the surface is
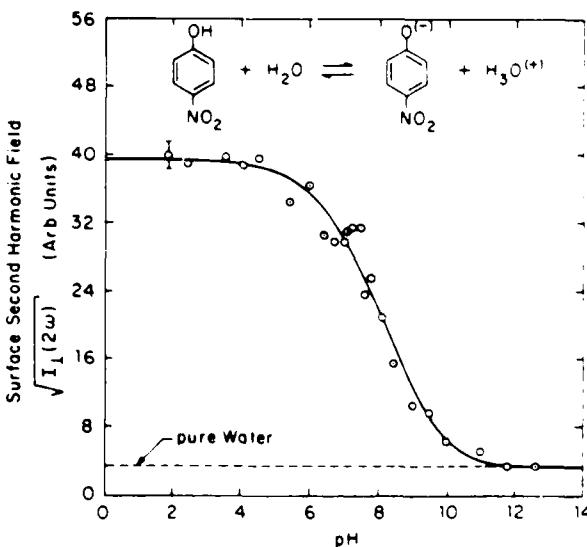

FIG. 1. s-polarized surface second harmonic field, $E_s^{2\omega} [\sqrt{I (2\omega)}]$ of an aqueous solution of p-nitrophenol as a function of pH at 22 °C.

also unlikely since its presence would have produced a second harmonic signal much above that of water. We therefore conclude that the anion once formed does not stay at the surface and thus in alkaline media the interfacial region is made up of water molecules alone. From our second harmonic measurements we estimate that the ratio of A^- to HA is at least 50–100 times smaller in the interfacial region than in the bulk solution. The equilibrium in the interfacial region is thus clearly shifted towards the neutral molecules and differs markedly from its bulk value.

Although Fig. 1 is similar to a neutralization curve the $pH(7.9 \pm 0.2)$ at which the SH signal amplitude reduces to half its maximum value is different from the $pH(7.15)$ at which the concentration of undissociated HA in the bulk decreases to half. This is not surprising since the bulk pH controls the concentration of HA in the bulk, while the surface concentration N_s is not a simple linear function of the latter. This point will be pursued in our future work. It is furthermore of interest to note the sensitivity of surface second harmonic generation as demonstrated by these results. Under the conditions described in this work, when the SH signal changes by a factor of 80, we found that the surface tension changed by less than 10%.

[1] Y. R. Shen, J. Vac. Sci. Technol. B 3, 1464 (1985).

[2] N. Bloembergen, R. K. Chang, S. S. Jha, and C. H. Lee, Phys. Rev. 174, 813 (1968).

[3] J. M. Chen, J. R. Bower, and C. S. Wang, Jpn. J. Appl. Phys. Suppl. 2, 711 (1974).

[4] P. Di Lazzaro, P. Mahatoni, and F. De Martini, Chem. Phys. Lett. 114, 103 (1985).

[5] G. Marowsky, A. Giersulski, and B. Dick, Opt. Commun. 52, 339 (1985), and references therein.

[6] Z. Chen, W. Chen, J. Zheng, W. Wang, and Z. Zhang, Opt. Commun. 54, 305 (1985).

[7] N. E. van Wyck, E. W. Koenig, J. D. Bynes, and W. M. Hetherington III, Chem. Phys. Lett. 122, 153 (1985).

[8] R. M. Corn, M. Romagnoli, and M. D. Levenson, J. Chem. Phys. 81, 4127 (1984).

[9] T. T. Chen, K. U. von Raben, D. V. Murphy, R. K. Chang, and B. L. Laube, Surf. Sci. 143, 369 (1984).

[10] G. L. Richmond, Langmuir 2, 132 (1986), and references therein.

[11] D. F. Voss, M. Nagumo, L. S. Goldberg, and K. A. Bunding, J. Phys. Chem. 90, 1834 (1986).

[12] J. M. Hicks, K. Kemnitz, K. B. Eisenthal, and T. F. Heinz, J. Phys. Chem. 90, 560 (1986).

[13] K. Kemnitz, K. Bhattacharyya, J. M. Hicks, J. R. Pinto, K. B. Eisenthal, and T. F. Heinz, Chem. Phys. Lett. 131, 285 (1986).

IR diode laser study of vibrational energy distribution in CO₂ produced by UV excimer laser photofragmentation of pyruvic acid[a)]

James A. O'Neill, Thomas G. Kreutz, and George W. Flynn

*Department of Chemistry and Columbia Radiation Laboratory, Columbia University, New York, New York 10027*

Time domain absorption spectroscopy using a tunable, infrared diode laser has been used to monitor the vibrational excitation of CO₂ produced in the 193 nm excimer laser photolysis of gas phase pyruvic acid (CH₃COCOOH). Nascent vibrational populations were measured in the following ten vibrational states of CO₂: 00^00, 01^10, 02^20, 03^30, 04^40, 00^01, 00^02, 00^03, 01^11, and 02^21. Approximately 97% of the CO₂ photoproduct is observed to be directly formed in the vibrational ground state. The remaining molecules are formed with a significant degree of vibrational excitation, having mode temperatures $T(\nu_2) = 1800 \pm 150$ K, $T(\nu_3) = 3700 \pm 1000$ K, $T(\nu_2 + \nu_3) = 2000 \pm 400$ K. The present experimental data suggest that the 193 nm photolysis may proceed through more than a single dissociation channel and involve a number of different photofragments.

INTRODUCTION

In recent years, the photodissociation dynamics of small molecules has become a subject of great interest.[1-5] Measurements of fragment energy distributions and angular orientations have yielded a wealth of state-specific experimental information about these dissociation processes. Such information can provide insight into the geometry of the dissociating species and the nature of the potential energy surfaces involved.[3-5] In addition, theoretical models involving "half-collision" arguments and molecular dynamics on potential energy surfaces have become quite refined. The vast majority of this experimental and theoretical work has sought to provide a detailed interpretation of the dissociation dynamics of predominantly diatomic and triatomic species, while the photolysis of most larger polyatomic molecules is less well understood.

Of particular importance in understanding the dynamics of photofragmentation processes are experimental studies of vibrational energy disposal in the photoproducts. Such experiments yield information about mode-specific excitation of photolysis products, which can in some cases be extrapolated back to provide a better understanding of the dissociation mechanism.[6-8] A variety of techniques, including infrared fluorescence, laser-induced fluorescence, and infrared laser absorption, have been employed to probe these product state energy distributions.[1] A new technique, which is presently being developed in our laboratory to investigate chemical dynamics and molecular energy transfer, utilizes a tunable diode laser absorption probe.[9-11] This technique has been employed to monitor the time resolved IR absorption of CO₂ molecules undergoing collisions with fast hydrogen and deuterium atoms as well as with electronically excited

mercury, NO₂, and azulene.[9-14] The high resolution (~0.0003 cm⁻¹) and the wide tunability of the diode laser make possible vibrationally and rotationally state-specific studies in a wide variety of molecules. In this experiment, virtually any rovibrational state of CO₂ can be probed by monitoring transitions of the type

$$CO_2(mn^lp, J) + h\nu(4.3 \ \mu m) \rightarrow CO_2(mn^lp + 1, J \pm 1) \tag{1}$$
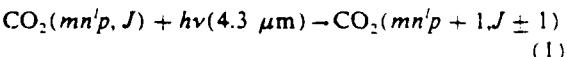
which share the large absorption coefficient of the ν_3 fundamental.[15] The letters m, n, p, and J are (respectively) the quantum numbers for the symmetric stretching (ν_1), bending (ν_2), and antisymmetric stretching (ν_3) vibrations, and the rotational motion of the molecule.

In the present experiments, the UV photolysis of pyruvic acid was investigated by monitoring the IR absorption of its CO₂ dissociation product. Previous photodissociation studies in both the gas and liquid phases have shown that pyruvic acid undergoes both thermal and photochemical decomposition to yield predominantly acetaldehyde and CO₂[16-21]:

$$CH_3COCOOH + h\nu(193 \ nm) \rightarrow CH_3CHO + CO_2(mn^lp). \tag{2}$$
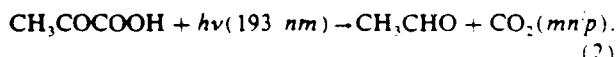
In one experiment which examined the vapor phase photolysis of pyruvic acid at 366 nm, small amounts (1%-2%) of methane and carbon monoxide were measured among the photolysis products.[19] Several mechanisms for the dissociation of this molecule have been proposed. One of these involves a four-center mechanism which leads to the direct production of acetaldehyde and CO₂, while another involves a five-center process which yields an hydroxycarbene intermediate species. A recent study which measured the 4.3 μm CO₂ IR fluorescence as a function of dissociation wavelength supports the five-center mechanism involving the hydroxycarbene intermediate.[20] In another experiment, measurements of the infrared multiphoton decomposition of pyruvic acid suggest that the dissociation process occurs by unimolecular decomposition in high vibrational levels of the

ground electronic state.[22] The UV photolysis of pyruvic acid is also thought to proceed through an internal conversion step where highly vibrationally excited ground state molecules decompose to form acetaldehyde and CO_2.[20]

In a previous study, we measured the relative production of CO_2 bending and antisymmetric stretching quanta formed in the 193 nm photolysis of pyruvic acid by probing the $01^10 \rightarrow 01^11$ and $00^01 \rightarrow 00^02$ transitions in CO_2.[21] The present experiments extend this work by measuring the complete distribution of vibrational energy deposited in the CO_2 bending, stretching, and combination state manifolds. We have examined such overtone levels as 02^20, 03^30, 04^40, 00^02, 00^03, 01^11, and 02^21, as well as the 00^00 ground state. In addition, we present preliminary information concerning the degree of rotational excitation in the CO_2 photofragment.

EXPERIMENTAL

The experimental apparatus employed in this study is similar to that used in several previous IR diode–UV excimer laser double resonance experiments carried out in our laboratory.[9-11,14] Most of the experimental details have been discussed previously, and only new features will be described here. The photolysis mixture, consisting of either neat pyruvic acid or a mixture of pyruvic acid and argon, flows through a cylindrical gas sample cell 207 cm long (see Fig. 1). The 4.3 μm cw output of a tunable IR diode laser (Laser Analytics) and the 193 nm pulsed output of a UV excimer laser (Lambda Physik) are collinearly propagated through the cell by the use of an IR transparent MgF_2 beam splitter coated for high reflectivity at 193 nm. The excimer laser is operated in an unstable resonator configuration which minimizes the divergence of the UV beam in the far field and allows for the propagation of the photolysis beam over the

long distances (5–7 m) necessary for the present experiments. Time-dependent changes in the absorption of the diode laser beam caused by CO_2 produced in the photolysis process are monitored with a liquid nitrogen cooled InSb detector and matched amplifier (300 ns response time). The signals are acquired with a Biomation 8100 transient recorder and signal averaged with a Nicolet 1170 averager.

Using a 207 cm cell at a CO_2 pressure of 10 Torr, where pressure broadening begins to make the direct spectral observation of weak absorption features difficult, the present apparatus is able to detect the presence of molecules in states whose ambient population at room temperature is about 10^{-6} of the ground state population. In order to locate and identify transitions corresponding to high lying vibrational levels which have extremely small ambient populations at room temperature, a portion ($\sim 8\%$) of the IR beam is split off and directed through a CO_2 discharge reference cell.[23] This cell consists of a high voltage dc discharge applied to a low pressure mixture of CO_2, N_2, and He. It provides a steady state, non-Boltzman population of highly vibrationally excited CO_2 with effective vibrational mode temperatures as high as 2900 K.[24] We present in Fig. 2 typical absorption spectra observed through the auxiliary cell with the discharge both on and off. Note that transitions from high lying vibrational levels such as $00^03 \rightarrow 00^04$ are readily observed when the discharge is applied. This is remarkable because there is less than 1 molecule/cm³ in a given rotational level of the 00^03 vibrational state for a 1 Torr sample at 300 K. Using the discharge cell, we have been able to probe CO_2 vibrational levels as high as 05^50 and 00^04. The reference cell is also used to actively stabilize the diode laser frequency onto a specific absorption line during the experiment. The reference signal is sent to a lock-in amplifier whose derivative output is fed back to the diode laser current controller. This procedure "locks" the laser frequency to the peak of any given CO_2 absorption line appearing in the discharge cell.
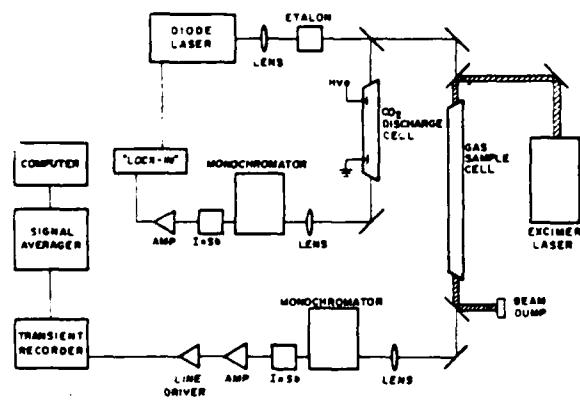

FIG. 1. A schematic diagram of the excimer/diode laser double resonance apparatus. The UV photolysis source is a pulsed excimer laser operating at 193 nm, and the IR probe source is a tunable diode laser emitting at approximately 4.3 μm. A portion ($\sim 8\%$) of the diode laser probe beam is diverted through a high voltage dc discharge cell used to locate the high vibrational states of CO_2 which are excited through an "anharmonic pumping" process. The discharge cell is also used as a reference cell in conjunction with a lock-in amplifier to frequency stabilize the diode laser. The IR probe beam is mode selected with a monochromator and detected with a liquid nitrogen cooled IR detector.
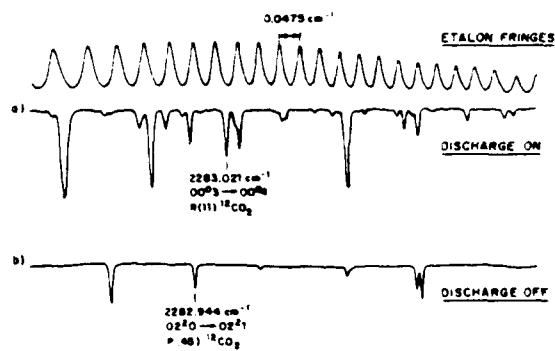

FIG. 2. Typical CO_2 absorption spectra using the CO_2 discharge reference cell. Top trace: Fringe pattern of a solid Ge etalon used for frequency referencing (free spectral range ~ 0.0475 cm⁻¹). Middle trace: CO_2 absorption spectrum in the vicinity of 2283 cm⁻¹ observed through the reference cell with the discharge applied (N_2:CO_2:He ~ 4:3:7; P_{tot} = 10 Torr; discharge current = 25 mA). Note the intensity of the absorption feature corresponding to the transition originating on the high lying 00^03 vibrational state. Lower trace: CO_2 absorption spectrum in the same spectral region with the discharge turned off.

In order to account for the shot to shot instabilities of our excimer laser photolysis source, the average pulse intensity is measured by diverting a portion of the UV beam to a Gentec power meter and accumulating the signals with a signal averager. The data is subsequently normalized by this factor.

The pyruvic acid used in these experiments (Aldrich 98%) was purified by repeated fractional distillation under partial vacuum conditions. Sample purity was checked with NMR spectroscopy. Pyruvic acid/argon samples were obtained by combining a concentrated mixture of pyruvic acid vapor in argon (made by flowing argon across a reservoir of the liquid acid) with a stream of pure argon. By varying the relative flow rates of the two streams, a variety of sample concentrations could be achieved while maintaining a total pressure of 5 Torr in the cell. The partial pressure of pyruvic acid in each mixture was determined by comparing the transient absorption signal of the mixture with signals obtained using known pressures of pure pyruvic acid in the cell. The sample flow speed was sufficiently rapid to prevent the buildup of photoproduct contaminants in the cell. It is particularly important to avoid the accumulation of vibrationally relaxed CO_2 molecules whose significant room temperature populations in the 00^00 and 01^10 levels can distort measurements of the nascent populations in these levels.

In a typical experiment, the change in the diode laser absorption $(mn^lpJ) \rightarrow (mn^lp + 1,J \pm 1)$ is measured following the UV photolysis of pyruvic acid/argon mixtures. Of particular interest is the prompt rise in signal intensity which immediately follows the excimer laser pulse (20 ns

width). The amplitude of this detector limited rise is a measure of the number of CO_2 molecules found in the (mn^lpJ) level after a time interval of 300 ns, the response time of the detection system. The sample pressure and composition are chosen specifically to influence the dynamics occurring within this time period. For example, the partial pressure of pyruvic acid is kept at 40–50 mTorr in order to assure adequate signal intensity while minimizing the partial pressure of the CO_2 photoproduct. This ensures that vibrational relaxation within each CO_2 vibrational mode (which occurs within 2–25 CO_2–CO_2 collisions) is insignificant within the rise time of the detector. In addition, excess argon (typically 5 Torr) is added to the pyruvic acid in order to ensure complete rotational relaxation in each vibrational level of the nascent CO_2 molecules within the rise time of the signal. As a result, the population within a particular vibrational level may, in principle, be obtained by measuring only one absorption line out of that level since the relative intensities of the absorption lines are given by the rotational Boltzman distribution. However, the vibrational populations reported in the next section are extracted from measurements taken on two or more absorption lines per vibrational state, each of which was found to yield similar results for the population of the entire vibrational level.

RESULTS

Immediately following the 193 nm photolysis of pyruvic acid, transient IR absorption signals were observed while monitoring the following CO_2 transitions:

$00^00 \rightarrow 00^01$ $P(60)$;	$00^01 \rightarrow 00^02$ $P(13),P(17),P(37)$;
$01^10 \rightarrow 01^11$ $P(49),P(35),P(30)$;	$00^02 \rightarrow 00^03$ $P(10),R(14),R(20)$;
$02^20 \rightarrow 02^21$ $P(37),P(22),P(17)$;	$00^03 \rightarrow 00^04$ $R(23),R(25)$;
$03^30 \rightarrow 03^31$ $P(24),P(8),P(33)$;	$01^11 \rightarrow 01^12$ $P(8),P(24)$;
$04^40 \rightarrow 04^41$ $P(10),P(20)$;	$02^21 \rightarrow 02^22$ $P(10)$;
$10^00 \rightarrow 10^01$ $P(48)$;	$00^00 \rightarrow 00^01$ $R(40),R(48)$ $^{13}CO_2$.

Our investigation focuses on the bending (v_2) and asymmetric stretch (v_3) vibrational modes of CO_2. These modes are only weakly coupled to each other, requiring approximately 30 000 CO_2–CO_2 collisions for intermode equilibration.[25,26] Rovibrational levels in the bending and symmetric stretch (v_1) modes, on the other hand, are strongly coupled via Fermi resonances and easily reach vibrational equilibration within the rise time of the detector. As a result, these experiments do not differentiate between the nascent populations formed in such states as 10^00, 02^00, and 02^20. Measurements of the vibrational populations for the 10^00 and 02^20 states yield a ratio of $N_{10^00}/N_{02^20} = 0.40 \pm 0.10$, which is in excellent agreement with the value of 0.39 predicted by Boltzmann statistics at room temperature. This indicates that the Fermi-mixed v_1 and v_2 vibrational states have attained thermal equilibrium through collisions with argon within the response time of the detector. Nevertheless, intramode vi-

brational equilibration (i.e., between 01^10, 02^20, 03^30, etc.) does not occur on this time scale.

The absorption signals obtained by our apparatus are a measure of the difference in population between the upper and lower levels of the IR transition. In order to determine the nascent population in a particular vibrational level (mn^lp), it was necessary to take into account the nascent population in the upper level $(mn^lp + 1)$. In the present experiments, contributions by the upper state populations to the signals from low lying vibrational levels such as 00^00, 01^10, 02^20, 00^01, and 00^02 were determined directly by probing (respectively) the 00^01, 01^11, 02^21, 00^02, and 00^03 states. For higher vibrational levels, it was necessary to estimate the contribution of the upper state by extrapolating from the measured data. Since the nascent populations were found to fall off rapidly with increasing energy, the upper state contributions were usually negligible.

$01^1 0 \longrightarrow 01^1 1$, P(30)

40 mTorr Pyruvic Acid
+ 5 Torr Argon

20 μsec full scale

$00^0 1 \longrightarrow 00^0 2$, P(37)

40 mTorr Pyruvic Acid
+ 5 Torr Argon

40 μsec full scale

$00^0 1 \longrightarrow 00^0 2$, P(37)

40 mTorr Pyruvic Acid
+ 5 Torr Argon

200 μsec full scale

Probe: $02^2 0 \longrightarrow 02^2 1$ P(22)

$P_{Pyruvic\ Acid}$ = 200 mtorr; λ = 193 nm

50 μsec full scale

A) 0 torr Ar

B) 1 torr Ar

C) 2 torr Ar

D) 3 torr Ar

FIG. 4. Transient absorption signals obtained while monitoring the $02^2 0 \rightarrow 02^2 1$ P(22) transition in CO_2 following the 193 nm photolysis of pyruvic acid. These curves illustrate the effect of added Ar on the rotational equilibration process which is responsible for feeding population into this state. As the pressure of Ar is increased, the initial change in absorption of the diode laser becomes more rapid because the rate of rotational equilibration is increased.

FIG. 3. Transient absorption signals obtained while monitoring the lowest bending and asymmetric stretching levels of the CO_2 photofragment. The prompt rise in the signal corresponds to the direct formation of CO_2 molecules in the rovibrational level being probed by the diode laser. The initial population of CO_2 found in the $00^0 1$ level is substantially smaller than that found in $01^1 0$, but the long time $00^0 1$ population increases dramatically due to collisional excitation.

Bending excitation in the CO₂ photoproduct

Transient absorption signals were obtained while monitoring the $01^1 0$, $02^2 0$, $03^3 0$, and $04^4 0$ bending mode vibrations in the CO_2 fragment. The $01^1 0$ signal, shown in Fig. 3, exhibits a detector limited rise (immediately following the excimer laser pulse) corresponding to the CO_2 molecules which are formed directly or relax into the $01^1 0 J = 30$ state. Following the prompt rise, we observe more complex temporal behavior which reflects the subsequent relaxation during collisions with other species present in the sample cell. Qualitatively similar temporal profiles were observed in all of the bending level signals.

Evidence of rotational excitation within the $02^2 0$ vibrational state is seen in the $02^2 0$ P(22) and P(37) signals (not shown) for pure pyruvic acid at pressures below 0.1 Torr. At pressures as low as 40 mTorr, the P(37) transition displays an initial rise which is detector limited. The P(22) absorption signal, on the other hand, is not detector limited and rises at a rate which is dependent on the total gas pressure. This suggests that CO_2 molecules which are produced in the $02^2 0$ pure bending state are also formed with significant rotational excitation. The lower rotational levels are filled during subsequent rotational relaxation. For states such as $02^2 0$, vibrational energy transfer from levels $10^0 0$, $02^0 0$ can also occur as a result of collisions as noted above. The effect of

increased argon pressure on the initial rise time of the $02^2 0$ P(22) absorption signal is shown in Fig. 4. The addition of 5 Torr of argon was found to ensure complete rotational relaxation within the rise time of the detector. Subsequent measurements of other CO_2 vibrational levels were carried out under these conditions.

In Fig. 5 we present the relative nascent vibrational populations measured for the $01^1 0$, $02^2 0$, $03^3 0$, and $04^4 0$ bending states of CO_2. The logarithm of the vibrational populations is seen to decrease linearly with the vibrational energy, suggesting a mode "temperature" for the bending levels which is significantly hotter than 300 K. By fitting a line to the levels with bending excitation, we find that the effective mode temperature is almost 1800 K. While mode temperatures normally include the ground state, our results (below) suggest that the vibrationally excited CO_2 photoproduct appears to be the result of a dissociation process which is different from that which forms ground state CO_2 molecules. As a result, the two populations must be treated separately.

Antisymmetric stretching excitation in the CO₂ photoproduct

Transient absorption signals were obtained while monitoring the $00^0 1$, $00^0 2$, and $00^0 3$ antisymmetric stretching states of CO_2 following the photodissociation pulse. The $00^0 1$ signal, shown in Fig. 3, displays a small, detector limited rise followed by a much slower, large amplitude increase. The fast rise is caused by CO_2 molecules which are directly produced with antisymmetric stretching vibration as a result of the photolysis process. The slower, large amplitude rise is due to energy transfer into the asymmetric stretch mode as a result of collisions with hot photofragments. The addition of
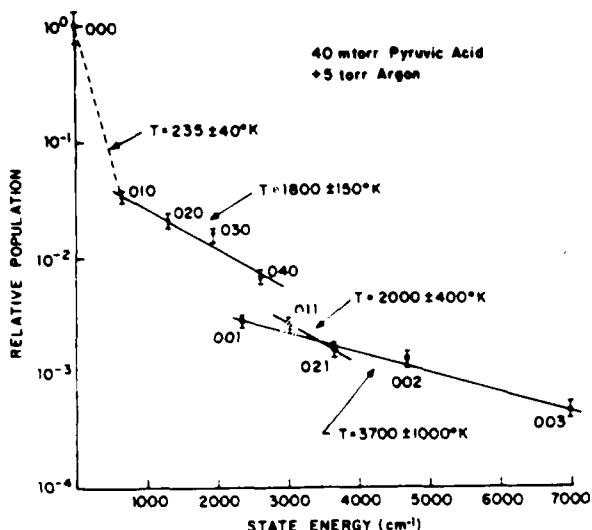
FIG. 5. Nascent distribution of population among the various vibrational states of the CO_2 photoproduct immediately following the photolysis pulse. Note the large abundance of molecules which are directly formed in the ground state as well as the hot mode temperatures which are suggested by the distribution of molecules among the vibrationally excited states of CO_2. The v_1 and v_2 modes in CO_2 are tightly coupled through the Fermi resonance pairs present in the molecule. Thus, with the large excess of Ar present in these experiments, it is not possible to obtain unambiguous results for the nascent populations in such levels as 10^00, 02^00, and 02^20 which are rapidly equilibrated by CO_2–Ar collisions during the rise time of the detection system.

argon to pure pyruvic acid caused no change in the amplitudes or rates of these features, suggesting that rotational relaxation does not play a significant role in these signals. The nascent vibrational population of each of the antisymmetric stretching states is shown in Fig. 5. Note that the absolute number of molecules directly produced in these levels is much smaller than that found in the vibrational states of the bending manifold, but the population distribution within the stretching mode is substantially hotter than the bending states, having an effective mode temperature of 3700 K.

Bend/stretch excitation in the CO_2 photoproduct

Excitation in bend/stretch combination states of the CO_2 photoproduct was observed by measuring the transient absorption signals for the $01^11 \rightarrow 01^12$ and $02^21 \rightarrow 02^2 2$ transitions. These signals demonstrate temporal behavior which is qualitatively similar to that shown by the pure antisymmetric stretching levels (see Fig. 3), showing an initial prompt rise followed by a slower, large amplitude increase due to vibrational energy transfer from hot photofragments. The nascent populations for the 01^11 and 02^21 level, shown in Fig. 5, have an effective vibrational temperature of approximately 2000 K. This value lies between the mode temperatures of the pure bending levels and the pure asymmetric stretch states, just as the combination state vibrational populations lie in between the pure bend and asymmetric stretch populations.

Formation of ground state CO_2 (00^00) photoproduct

The production of CO_2 in the vibrational ground state was measured by probing the $00^00 \rightarrow 00^01$ transition. Since the amount of CO_2 present in the atmosphere is sufficient to cause complete absorption of the diode laser beam at frequencies near the peak of the room temperature rotational distribution, we found it necessary to monitor the $00^00 \rightarrow 00^01$ transitions of $^{13}CO_2$ during the photolysis. The small (1%) natural abundance of this species limits atmospheric absorption of the diode laser and allows examination of rotational levels close to the peak of the room temperature Boltzmann distribution. We assume that the dynamics of the dissociation process involving $^{13}CO_2$ are the same as those involving the lighter $^{12}CO_2$ species. In theory, the vibrational population can also be measured by monitoring a very high rotational line in the ground state of $^{12}CO_2$. However, this latter method involves the extraction of an entire vibrational population from a weak absorption line located on the wings of the rotational distribution, and as a result, involves a great deal of experimental uncertainty.

The $^{13}CO_2$ $00^00 \rightarrow 00^01$ absorption signals exhibit an initial prompt rise which has a very large amplitude. Relative population measurements show that the number of molecules directly formed in this state during pyruvic acid photolysis is 30 ± 6 times that produced in the 01^10 bending state (see Fig. 5). In other words, 95%–98% of all CO_2 photofragments are formed in the ground vibrational level. This result is in excellent agreement with our previous study which showed (by absolute population measurements rather than the relative measurements presented here) that only 3% of the nascent CO_2 molecules are found in the 01^10 level.[21] In addition, an initial prompt rise (which was subsequently relaxed by the presence of excess argon) was seen in rotational levels as high as $^{12}CO_2$ $00^00 J = 60$. This suggests a high degree of rotational excitation in CO_2 molecules which are produced in the ground state during the photolysis.

DISCUSSION

There are several interesting features of the pyruvic acid photodissociation process which can be compared to the data obtained in the present experimental study. Figure 6 shows an energy diagram for the photofragmentation event. Overall, the production of carbon dioxide and acetaldehyde from pyruvic acid is exothermic by 5 kcal/mol. This energy (~ 1750 cm^{-1}) is quite small compared to that available from 193 nm photolysis (152 kcal/mol or 53 150 cm^{-1}) because of the high energy of the UV photon. Even if the products initially produced by photodissociation are the hydroxycarbene intermediate and carbon dioxide, there is still a huge amount of energy left over (20 650 cm^{-1}) to be divided among the fragments. The present data, which show that more than 95% of the CO_2 molecules are produced in the vibrational ground state, leave this rather large energy unaccounted for.

Simple geometric considerations suggest that the CO_2 molecules produced in the photofragmentation process should have large amounts of bending energy. The CO_2 cen-
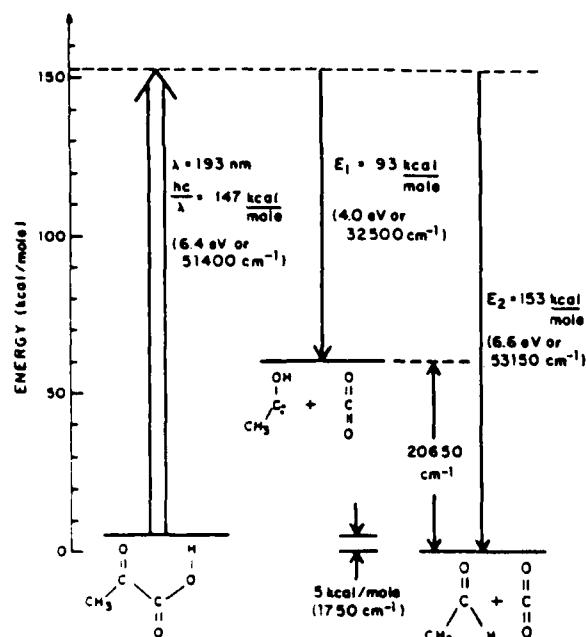
FIG. 6. An energy level diagram indicating the energetics of the 193 nm photolysis of pyruvic acid. The final CO_2 and acetaldehyde products are exothermic by 5 kcal/mol. The formation of a transient hydroxycarbene intermediate species along the dissociation pathway requires approximately 54 kcal/mol of the total 147 kcal/mol which is available in the photolysis process.

ter in the acid is sp^2 hybridized with a 120 deg angle for the O–C–O bond, while free CO_2 has linear, sp hybridization. Assuming that the H atom on the acid center is transferred to the CH_3CO fragment without significant rearrangement during photodissociation, a simple Franck–Condon argument would predict a large amount of bending excitation in the final CO_2 product due to the changing geometry alone. Clearly, some molecules are produced with substantial bending excitation since ~3% of the CO_2 is found in the $01^10, 02^20, 03^30, 04^40$ states with a distribution corresponding to a rather high temperature (1800 K). Nevertheless, this represents an extremely small fraction of the total energy held by the photodissociated species. Antisymmetric stretching excitation might also be expected as the C–O bond lengths change during the evolution from pyruvic acid to free CO_2, but the amount of such excitation is observed to be an order of magnitude less than that found in the bending mode. Of course, considerable rearrangement of the pyruvic acid would be expected to occur through a five center transition state in which the acid H atom is shared by two O atoms which are in turn attached to two different carbon atoms. In such a structure the simple geometric arguments presented above would no longer be expected to hold because the dissociating species would not necessarily have the ground state geometry.

The majority of the missing energy is unlikely to have been carried away by a vibrationally hot acetaldehyde fragment. While acetaldehyde does have 15 vibrational modes compared to 4 for CO_2, it is extremely difficult to see how such a vibrationally hot species ($20\,650–53\,150$ cm^{-1}) could separate from CO_2 without leaving the latter molecule

much hotter than is observed, especially considering that a substantial number of the acetaldehyde modes have higher vibrational frequencies than those of CO_2.

There appear to be a variety of possibilities which can account for the "missing energy." First, there is more than enough energy available from the very energetic 193 nm UV photon to break several bonds in pyruvic acid. Thus, photodissociation may actually produce more than two fragments (e.g., CO_2, CO, CH_4, etc.), and some of the missing energy could simply be going into molecular bond breaking. Small amounts of CO and CH_4 have been observed during the 366 nm of pyruvic acid, but a product analysis study of this photoreaction has not been made at 193 nm.[19] We are presently investigating this point by using diode lasers to probe for the presence of CO and CH_4 following 193 nm photodissociation of pyruvic acid. Preliminary results indicate that a measurable amount of CO is formed immediately upon photodissociation.

Second, the missing energy could be accounted for if electronically excited states of the products are formed. The most obvious of these is the hydroxycarbene intermediate/isomer which is $20\,650$ cm^{-1} above the ground state of normal acetaldehyde. The hydroxycarbene intermediate is especially appealing since pyruvic acid is expected to be an internally hydrogen-bonded monomer in the gas phase.[27] Nevertheless, if this is the only excited species formed, large amounts of energy are still available for vibrational excitation in the CO_2 product. A higher electronic state of acetaldehyde, the hydroxycarbene, or even CO_2 could, of course, account for a large amount of energy. In this regard, it is interesting to note that strong excitation of the CO_2 v_3 mode, which might result from the relaxation of such a highly excited species, was observed on a long time scale in the present experiments. In addition, absorption signals for the $01^10 \rightarrow 01^11$ transition in CO_2 show a similar long time filling of the fundamental bending state. While a full characterization of the long time behavior of all vibrationally excited states is needed, the present observations suggest that both bending and stretching levels of the product CO_2 molecule are populated to a similar extent on a long time scale, probably due to collisions with hot photofragments present in the cell after the dissociation process.

Finally, nascent vibrational populations have been measured in only ten low lying vibrational levels of CO_2. CO_2 could be produced in vibrational levels higher than those probed, i.e., the observed decrease in vibrational population with increasing energy may be part of a bimodal vibrational energy distribution. Mitigating against such a picture are "harvesting" experiments, where pyruvic acid was photodissociated in the presence of a large background of cold CO_2, which did not reveal the presence of vibrational energy held in higher vibrational levels.[31] Such experiments are not foolproof, however, and the possibility of a bimodal distribution cannot be entirely ruled out at the present time.

The possibility that a large amount of the initially deposited photon energy appears in the translational degrees of freedom of the recoiling fragments was considered by taking advantage of the high resolution of the diode laser and measuring the product CO_2 absorption linewidth immediately

following the photolysis event. Such measurements were performed for various $02^20 \rightarrow 02^21\ P, Q$, and R branch transitions at a pressure of 25 mTorr (pure pyruvic acid) where the time interval between collisions is $\sim 4\ \mu s$. Under these conditions, no linewidth changes were observed in the time interval approximately 1 μs after the excimer laser pulse. This observation suggests that little translational energy appears among product CO_2 molecules which are vibrationally excited during the photolysis process in the 02^20 state. The possibility that the large amount of ground state CO_2 photofragment is produced translationally hot is presently under investigation by measuring the linewidth of the $00^00 \rightarrow 00^01$ absorption transition after the photolysis process.

The presence of a large amount of energy in the translational degrees of freedom of the product molecules could also manifest itself in the rotational equilibration process which occurs in these experiments since translations and rotations equilibrate rapidly. The importance of such an effect, however, is expected to be small since under typical experimental conditions the temperature rise in the cell is $\leq 10°$. Such a small increase in temperature would not cause a large enough shift in the rotational distribution to account for the transient population changes which are observed in the present experiments.

Overall, the data taken in the present experiments may well indicate a dual channel mechanism for photodissociation. One channel appears to produce a small amount of CO_2 with a high degree of vibrational excitation, and could represent the conventional Franck–Condon picture in which the CO_2 center is changing from highly bent, sp^2 character to linear sp character. The second channel, which produces primarily ground state CO_2, must necessarily proceed through high energy transient species or multiple bond breaking. Given the high energy of the UV photon used in the present experiments, several surface crossings to products could be available during photofragmentation. Experiments performed at a variety of photolysis wavelengths would be very useful in helping to sort out the photodissociation mechanism.

Finally, the present experiments also indicate that CO_2 is produced with considerable rotational excitation. Measurements of rotational energy distributions in the 00^00, 01^10, and 02^20 vibrational levels are currently in progress.

CONCLUSIONS

A tunable IR diode laser absorption probe technique has been employed to examine nascent vibrational population distributions in the CO_2 photoproduct resulting from the 193 nm laser photolysis of pyruvic acid $CH_3COCOOH$. Approximately 97% of the CO_2 product is formed directly in the vibrational ground state, while the remaining CO_2 molecules are produced with a significant degree of vibrational excitation. These latter molecules exhibit vibrational energy distributions which correspond to the following mode temperatures:

$$T(v_2) = 1800 \pm 150\ K,$$

$$T(v_3) = 3700 \pm 1000\ K,$$

$$T(v_2 + v_3) = 2000 \pm 400\ K.$$

These results suggest that the 193 nm photolysis may proceed through two dissociation channels having markedly different fragmentation mechanisms. In addition, evidence of substantial rotational excitation has been observed.

The vibrational excitation in the CO_2 photofragment accounts for only a small fraction of the energy available in the dissociation process. The remaining 6 eV of available energy is sufficient for the production of electronically excited acetaldehyde or CO_2 products as well as the further fragmentation of the acetaldehyde molecule. Preliminary results indicate that CO molecules are also produced during the 193 nm photodissociation. A more complete picture of the photofragmentation dynamics should emerge with a wavelength dependent analysis of the dissociation products.

These experiments demonstrate the power of time domain diode laser absorption spectroscopy for the measurement of nascent vibrational state distributions in polyatomic molecules. Recent experiments at lower pressures have demonstrated the usefulness of this technique in extracting rotationally resolved product state distributions for a variety of experiments.[28] We are presently applying these methods to examine the nascent rotational distributions in the CO_2 photofragment from pyruvic acid in order to futher elucidate the dynamics of the 193 nm photolysis. The extremely high resolution of the diode laser ($\sim 0.0003\ cm^{-1}$) can also be used to measure the linewidth of the nascent product states in order to obtain information about the recoil velocity of the CO_2 fragment.

[1]S. R. Leone, Adv. Chem. Phys. **50**, 255 (1982).
[2]R. Bersohn, J. Phys. Chem. **88**, 5145 (1984).
[3]W. M. Gelbart, Annu. Rev. Phys. Chem. **28**, 323 (1982).
[4]K. F. Freed and Y. B. Band, in *Excited States*, edited by E. L. Lim (Academic, New York, 1977), Vol. 3 , p. 110.
[5]M. Shapiro and R. Bersohn, Annu. Rev. Phys. Chem. **33**, 409 (1982)
[6]W. H. Hermann and S. R. Leone, J. Chem. Phys. **76**, 4759 (1982).
[7]W. H. Hermann and S. R. Leone, J. Chem. Phys. **76**, 4766 (1982).
[8]S. L. Baughcum and S. R. Leone, Chem. Phys. Lett. **89**, 183 (1982).
[9]J. O. Chu, C. F. Wood, G. W. Flynn, and R. E. Weston, Jr., J. Chem. Phys. **80**, 1703 (1984).
[10]J. O. Chu, C. F. Wood, G. W. Flynn, and R. E. Weston, Jr., J. Chem. Phys. **81**, 5533 (1986).
[11]J. A. O'Neill, J. Y. Cai, G. W. Flynn, and R. E. Weston Jr., J. Chem. Phys. **84**, 50 (1986).
[12]B. B. Brady, G. B. Spector, L. Chia, and G. W. Flynn, J. Chem. Phys. **86**, 3245 (1987).
[13]B. B. Brady, L. Chia, and G. W. Flynn (work in progress)
[14]W. Jalenak, R. E. Weston, Jr., T. J. Sears, and G. W. Flynn, J. Chem. Phys. **83**, 6049 (1985).
[15]Air Force Cambridge Research Laboratories, AFCRL Atmospheric Absorption Line Parameters Compilation, Environmental Research Papers, No. 434, L.G. Hanscom Field, Bedford, MA, January 1973
[16]M. Urion, Ann. Chem. **1**, 5 (1934).
[17]J. F. Arnett, D. B. Larson, and S. P. McGlynn, J. Am. Chem. Soc. **95**, 7599 (1973).
[18]P. A. Leermakers and G. F. Versley, J. Am. Chem. Soc. **85**, 3776 (1963).
[19]G. F. Versley and P. A. Leermakers, J. Phys. Chem. **68**, 2364 (1964).

[20]R. N. Rosenfeld and B. Weiner, J. Am. Chem. Soc. 105, 3485 (1983).

[21]C. F. Wood, J. A. O'Neill, and G. W. Flynn, Chem. Phys. Lett. 109, 317 (1984).

[22]J. L. Beuchele, E. Weitz, and F. D. Lewis, Chem. Phys. Lett. 77, 280 (1981).

[23]T. G. Kreutz, J. A. O'Neill, and G. W. Flynn, J. Phys. Chem. (accepted).

[24]C. Dang, J. Reid, and B. K. Garside, Appl. Phys. B 27, 145 (1982).

[25]W. A. Rosser, Jr., A. D. Wood, and E. T. Gerry, J. Chem. Phys. 50, 4996 (1969).

[26]L. O. Hocker, M. A. Kovacks, C. K. Rhodes, G. W. Flynn, and A. Javan, Phys. Rev. Lett. 17, 233 (1966).

[27]W. J. Ray, J. E. Katon, and D. B. Phillips, J. Mol. Struct. 74, 75 (1981).

[28]J. A. O'Neill, C. X. Wang, J. Y. Cai, G. W. Flynn, and R. E. Weston, J. Chem. Phys. 85, 4195 (1986).

Rotationally resolved isotope effect in the hot atom collisional excitation of CO$_2$ (00°1) by time-dependent diode laser spectroscopy

Scott A. Hewitt[a] John F. Hershberger, and George W. Flynn
Department of Chemistry and Columbia Radiation Laboratory, Columbia University, New York, New York 10027

Ralph E. Weston, Jr.
Department of Chemistry, Brookhaven National Laboratory, Upton, New York 11973

The study of collisional energy transfer between translational and internal degrees of freedom has received increased attention in recent years.[1-8] An efficient source of translational energy is the excimer laser photolysis of small molecules to produce hot atoms with energies of several eV. In the present work, hot deuterium atoms have been produced by photolysis of D$_2$S:

$$D_2S + h\nu(193 \text{ nm}) \rightarrow D^*(E_{tr} = 2.16 \text{ eV}) + DS. \quad (1)$$
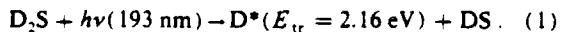
Inelastic scattering then produces excited CO$_2$:

$$D^* + CO_2 \ (00^00,J) \rightarrow D + CO_2 \ (00^01, J'). \quad (2)$$
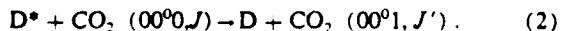
High-resolution diode laser spectroscopy is then used to probe the rotational level of the CO$_2$ (00°1) state:

$$CO_2 \ (00^01, J') + h\nu(4.3 \ \mu m) \rightarrow CO_2 \ (00^02, J' - 1). \quad (3)$$
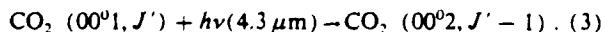
In this Communication we report the observation of a preference for excitation of high J relative to low J levels of CO$_2$ (00°1) by hot D atoms compared to hot H atoms.

This experiment is based on an earlier study of the rotational distribution of CO$_2$ (00°1, J) due to collisions with hot H atoms at 2.30 eV produced by the photolysis of H$_2$S.[9] In that study, the rotational distribution was found to be peaked at $J = 33$, substantially higher than the room-temperature Boltzman distribution peak of $J = 15$. In work completed prior to that study,[7,8] the total number of quantum in the CO$_2$ asymmetric stretch and bend vibrations was measured, and the data indicated a substantial isotope effect with the heavier D* atom exciting the lower frequency bend preferentially.

In the present experiments, 193 nm pulses from an ArF excimer laser with unstable resonator optics (Lambda-Physik EMG 201) are propagated down a 9 ft sample cell through which flows a 1/1 mixture of D$_2$S/CO$_2$ or H$_2$S/CO$_2$. Tunable cw radiation from a diode laser (Laser Analytics) at 4.3 μm is propagated through the cell collinearly with the excimer beam. The diode laser is tuned to specific rotational lines of the 00°1 → 00°2 vibrational band of CO$_2$. Time-resolved changes in the transmitted IR intensity are measured with an InSb detector. The signals are digitized, summed, and sent to a computer for storage and later analysis.

Transient absorption signals of the 00°1 → 00°2 $P(13)$, $P(19)$, $P(29)$, $P(35)$, $P(39)$, $P(41)$, $P(51)$, $P(59)$, and $P(65)$ transitions in CO$_2$ were obtained at sample pressures of 25 mTorr. The low pressures used assure that rotational relaxation of the CO$_2$ ($\sim$4 μs) is substantially slower than the rise time of the detection system ($\sim$0.5 μs), permitting direct measurement of nascent rotational populations. The observed signals for high rotational levels of 00°1 CO$_2$ have an initial fast rise due to direct formation of excited CO$_2$ by

RATIO OF $CO_2(00^\circ 1, J)$ EXCITATION BY H^* vs. D^*

$H_2S/CO_2 = 1/1$; 25 mtorr
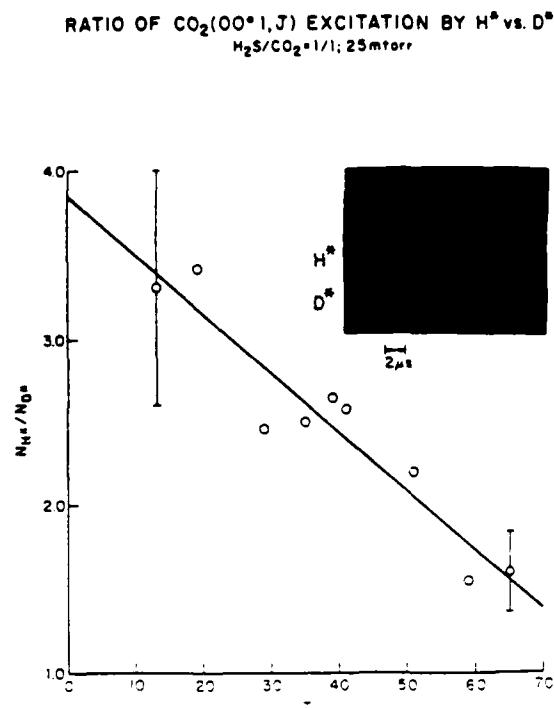
FIG. 1. Experimental values for N_{H^*}/N_{D^*} vs CO_2 rotational level (J) for 1/1 mixtures of H_2S/CO_2 and D_2S/CO_2 at 25 mTorr. $N_{H^*}(N_{D^*})$ is the population increase in a given $00^\circ 1$ rotational state after the $H^*(D^*)/CO_2$ collisional interaction. The solid curve is a linear least-squares fit of the data. The inset scope trace shows the change in absorption of the diode probe beam for the $00^\circ 1 - 00^\circ 2$ $P(39)$ transition after 193 nm photolysis of H_2S in CO_2 and D_2S in CO_2 in a 9 ft cell.

hot H or D atoms (see inset of Fig. 1), followed by a slower decay which is attributed to rotational relaxation. For low rotational levels of $00^\circ 1$ CO_2, the fast rise is small or nonexistent, and the signals have a slow rise due to rotational relaxation which fills the lower J states.

To obtain the rotational distribution directly, the fast rise amplitudes must be normalized for excimer and diode laser intensities and absorption line strengths; this was done in the earlier H^* atom experiment.[9] In the present experiments, to avoid the errors involved with such normaliza-

tions, ratios of H^* and D^* atom fast rise amplitudes were measured by locking the diode laser on a given $00^\circ 1$, $J' - 00^\circ 2$,$J' - 1$ transition and flowing through the gas cell an H_2S/CO_2 sample and then a D_2S/CO_2 sample. When plotted as a function of rotational level, these ratios show a significant negative slope (see Fig. 1). From the ratios and the previously measured H^* atom distribution,[9] the D^* atom distribution can be obtained. While the overall vibrational excitation of CO_2 $(00^\circ 1, J)$ by H^* atoms is about twice as efficient as excitation by D^* atoms, Fig. 1 shows a clear preference for higher J by deuterium as compared to hydrogen.

The available angular momentum in a hot H or D collision is $L = \mu bg$ where μ is the reduced mass, b the impact parameter, and g the relative speed. For fixed b, D^*/CO_2 collisions have a $\sqrt{2}$ larger L than H^*/CO_2 collisions. This probably accounts for the increasing excitation probability of high J $00^\circ 1$ levels by D^* atoms compared to H^* atoms.

*) AT&T Bell Laboratories Ph.D. Scholar

[1] C. R. Quick, Jr., R. E. Weston, Jr., and G. W. Flynn. Chem. Phys. Lett. 83, 15 (1981).

[2] F. Magnotta, D. J. Nesbitt, and S. R. Leone. Chem. Phys. Lett. 83, 21 (1981).

[3] C. F. Wood, G. W. Flynn, and R. E. Weston, Jr. J. Chem. Phys. 77, 4776 (1982).

[4] C. A. Wight and S. R. Leone. J. Chem. Phys. 78, 4875 (1983).

[5] J. O. Chu, G. W. Flynn, and R. E. Weston, Jr. J. Chem. Phys. 78, 2990 (1983).

[6] C. A. Wight and S. R. Leone. J. Chem. Phys. 79, 4823 (1983).

[7] J. O. Chu, C. F. Wood, G. W. Flynn, and R. E. Weston, Jr. J. Chem. Phys. 81, 5533 (1984).

[8] J. A. O'Neill, J. Y. Cai, G. W. Flynn, and R. E. Weston, Jr. J. Chem. Phys. 84, 50 (1986).

[9] J. A. O'Neill, C. X. Wang, J. Y. Cai, G. W. Flynn, and R. E. Weston, Jr. J. Chem. Phys. 85, 4195 (1986).

Diode Laser Absorption Probe of Vibration–Vibration Energy Transfer in CO$_2$

Thomas G. Kreutz, James A. O'Neill, and George W. Flynn*

Department of Chemistry and Columbia Radiation Laboratory, Columbia University, New York, New York 10027 (Received: August 5, 1987)

Time domain absorption spectroscopy with a tunable diode laser has been used to measure the rate constant for vibration-to-vibration energy transfer in CO$_2$ via the near-resonant "collisional up-pumping" process CO$_2$(00°1) + CO$_2$(00°1) $\rightleftharpoons$ CO$_2$(00°2) + CO$_2$(00°0). The diode laser, in combination with a CO$_2$ discharge reference cell, is used to monitor the time-dependent population increase within the CO$_2$(00°2) state in a 10:1 mixture of Ar and CO$_2$ following pulsed excitation of the CO$_2$(00°1) level by a Q-switched CO$_2$ laser. The intramode relaxation rate constant, k, for this process is found to be $(5.6 \pm 0.7) \times 10^6$ s^{-1} Torr^{-1} or $(1.7 \pm 0.2) \times 10^{-10}$ cm^3 s^{-1} molecule^{-1} at 298 K.

Introduction

The dynamics of both pulsed and CW CO$_2$ lasers have been the subject of extensive research for many years [1-3] The behavior of these systems is now rather well understood, and kinetic models agree well with experimental results. Such models are based upon numerous rate constants for vibration-to-vibration (V–V) and vibration-to-translation (V–T) energy transfer between CO$_2$, N$_2$, and rare gases, values for which have been measured by infrared fluorescence and CO$_2$ laser double-resonance techniques. [2-10]

Perhaps the most fundamental V–V energy-transfer process in CO$_2$ is the near-resonant "collisional up-pumping" process

$$CO_2(00^01) + CO_2(00^01) \underset{k_{r}}{\overset{k}{\rightleftharpoons}} CO_2(00^02) + CO_2(00^00) \qquad \Delta\epsilon = -25 \text{ cm}^- \quad (1)$$
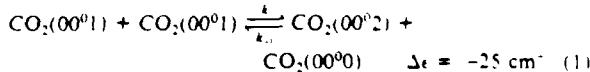
which serves as a model for how vibrational energy is distributed within the asymmetric stretch (ν_3) mode of CO$_2$. The intramode relaxation rate constant, k, for this process is expected to be approximately gas kinetic and has been calculated by Pack,[11] but an experimental value for the room-temperature rate constant has not been published to date.[10] We have recently measured this rate using the powerful combination of a tunable diode laser (TDL) spectrometer and a CO$_2$ discharge reference cell. The diode laser provides a high-resolution probe ($\sim$0.0003 cm^{-1}) for time domain absorption spectroscopy and has been recently used to probe CO$_2$ laser dynamics.[3] The CO$_2$ discharge reference cell

(1) Patel, C. K. N. *Phys. Rev. Lett.* **1964**, *12*, 588. **1964**, *13*, 617. *Appl. Phys. Lett.* **1965**, *7*, 290.

(2) Gordietz, B. F., Sobolev, N. N., Sokovikov, V. V., Shelepin, L. A. *IEEE J. Quantum Electron.* **1968**, *4*, 796. Moore, C. B., Wood, R. E., Hu, B. E., Yardley, J. T. *J. Chem. Phys.* **1967**, *46*, 4222.

(3) Dang, C., Reid, J., Garside, B. K. *Appl. Phys.* **1983**, *B31*, 163. *IEEE J. Quantum Electron.* **1983**, *19*, 755. Brimacombe, R. K., Reid, J., Znotins, T. A. *Appl. Phys.* **1985**, *B36*, 115.

(4) Stephenson, J. C., Moore, C. B. *J. Chem. Phys.* **1972**, *56*, 1295. Stephenson, J. C., Wood, R. E., Moore, C. B. *J. Chem. Phys.* **1971**, *54*, 3097. Yardley, J. T., Moore, C. B. *J. Chem. Phys.* **1967**, *46*, 4491.

(5) Rosser, Jr., W. A., Gerry, A. D., Wood, E. T. *J. Chem. Phys.* **1969**, *50*, 4996. Rosser, Jr., W. A., Sharma, R. D., Wood, E. T. *J. Chem. Phys.* **1971**, *54*, 1196. Rosser, Jr., W. A., Gerry, A. D. *J. Chem. Phys.* **1971**, *54*, 4131.

(6) Huddleston, R. K., Weitz, E. *Chem. Phys. Lett.* **1981**, *83*, 174.

(7) Temple, T. A., Suhre, D. R., Coleman, P. D. *Appl. Phys. Lett.* **1973**, *22*, 349. Stark, Jr., E. E. *Appl. Phys. Lett.* **1973**, *23*, 335. Jacobs, R. R., Pettipiece, K. J., Thomas, S. J. *Phys. Rev. A* **1975**, *11*, 54.

(8) Finzi, J., Moore, C. B. *J. Chem. Phys.* **1975**, *63*, 2285.

(9) Burak, I., Noter, Y., Szoke, A. *IEEE J. Quantum Electron.* **1973**, *9*, 541.

(10) Thomason, M. D. Ph.D. Thesis, University of Virginia. Los Alamos National Laboratory, LA9420-T, 1982.

(11) Pack, R. T. *J. Chem. Phys.* **1980**, *72*, 6140. Note that Pack's rate constant, k, written in terms of the disappearance of CO$_2$(00°1), is twice as large as that defined in eq 1 and 2.

COLLISIONAL UP-PUMPING

$$CO_2(00^\circ1) + CO_2(00^\circ1) \rightleftharpoons CO_2(00^\circ2) + CO_2(00^\circ0)$$

DIODE LASER ABSORPTION

$$CO_2(00^\circ2; J = 20) + \quad 4.3\mu m \longrightarrow CO_2(00^\circ3; J = 21)$$
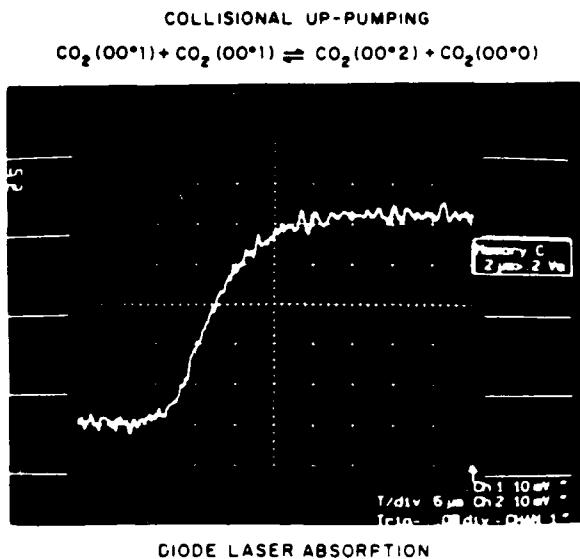
Figure 1. Time domain absorption signal for the $(00^02) \rightarrow (00^03)$ R(20) transition of CO_2 following excitation of the (00^01) level by 9.6-μm radiation from a Q-switched CO_2 laser. The sample is a 1:10 mixture of CO_2:Ar at a total pressure of 1.0 Torr and a temperature of 298 K. The trace represents the average of 32 000 signals and has a time base of 20 μs full scale

creates steady-state populations of vibrationally hot CO_2 and provides frequency references for thousands of high-lying rovibrational lines which are normally inaccessible at room temperature. The TDL/discharge cell combination allows virtually every important rovibrational level in the CO_2 laser to be probed with extremely high temporal and spectral resolution.

Experimental Section

In the present double-resonance experiment, the 9.6-μm output of a Q-switched CO_2 laser is propagated through a 2-m-long sample cell containing a 1:10 mixture of CO_2 and argon, populating the $CO_2(00^01)$ vibrational state. The argon promotes rotational relaxation within each CO_2 vibrational level without significantly affecting vibrational relaxation, so that the rotational levels are described by a Boltzmann distribution throughout the experiment. In addition, the rare gas limits radial diffusion of excited CO_2 molecules out of the probe beam. The diode laser, which is used to monitor (via absorption at 4.3 μm) the time-dependent populations in the (00^02) level, is copropagated along the cell axis, passed through a monochromator (to discriminate against competing spatial and longitudinal modes), and detected with a cooled (77 K) InSb detector. Time-resolved changes in the transmitted intensity of the diode radiation are acquired with a Biomation 8100 transient digitizer and averaged on a Nicolet 1170 signal averager. A trigger is provided by detecting a portion of the CO_2 laser pulse (~ 0.4-μs width) with a cooled (77 K) HgCdTe detector having a response time of ~ 0.1 μs. A typical time domain absorption signal for the (00^02) level is shown in Figure 1, where the diode laser is tuned to the $(00^02) \rightarrow (00^03)$ R(20) transition at 2313.959 cm^{-1}. In order to locate this absorption line and lock the diode laser to this frequency, a portion of the beam is split off before the sample cell and directed through a CO_2 discharge reference cell and into a monochromator/IR detector. The reference cell absorption signal is sent to a lock-in amplifier whose output is fed back into the diode laser current controller for frequency stabilization. This configuration fixes the frequency of the diode laser at the peak of a single absorption line throughout the duration of the experiment.

The discharge reference cell consists of a high-voltage dc discharge (16 kV; 25 mA) applied to a low-pressure mixture of CO_2, N_2, and He. It provides a steady-state, non-Boltzmann

CO₂ DISCHARGE CELL ABSORPTION SPECTRUM
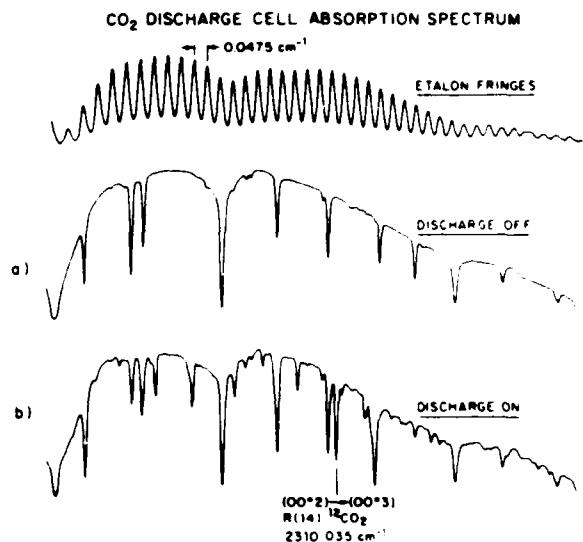
Figure 2. Typical CO_2 discharge cell absorption spectra in the 2300-cm region with the discharge (a) turned off and (b) ignited (25 mA). Note that the intensity of the $(00^02) \rightarrow (00^03)$ R(14) line at 2310.035 cm^{-1} is increased by 8 orders of magnitude when the discharge is turned on.

population of highly vibrationally excited CO_2, with "effective" (Treanor-type[12]) vibrational mode temperatures as high as 2900 K.[13] In this experiment, the discharge enhances the (00^02) population by more than 8 orders of magnitude (from $\sim 10^{-10}$ to 0.05 of the ground-state population), providing a precise frequency reference for each $(00^02) \rightarrow (00^03)$ absorption line. Accurate assignments of the additional lines present in the discharge are made possible by high-resolution FTIR spectroscopy of discharge excited CO_2.[14] Typical reference cell absorption spectra are shown in Figure 2. Note the dramatic increase of available absorption lines when the discharge is employed.

Data Analysis

The kinetics of the collisional up-pumping process given in eq 1 may be modeled simply by

$$N_2(t) = k[N_1(t)]^2 - k_{-1}N_0N_2(t) \tag{2}$$

where $N_n(t)$ is the population in (00^0n) and k and k_{-1} are the forward and backward rate constants. Since N_0 and $N_1(t)$ are not significantly perturbed by the up-pumping process, eq 2 is uncoupled from the analogous equations for N_0 and $N_1(t)$. As a result, the general solution to eq 2 may be written

$$N_2(t) = k \int_0^t [N_1(\tau)]^2 e^{-N_0 k_{-1}(t-\tau)} d\tau \tag{3}$$

As a first approximation, we assume that the (00^01) population has no significant time dependence on the time scale of interest, $1/N_0 k_{-1}$, and so $N_1(t)$ may be approximated by using a step function with amplitude N_1. Assuming $N_2(t<0) = 0$, the solution to eq 3 is given by

$$N_2(t) = N_2(\infty)[1 - e^{-N_0 k_{-1} t}] \tag{4}$$

where $N_2(\infty) = [N_1]^2 k / N_0 k_{-1}$.

Time domain absorption signals were taken on the (00^02) R(14) and R(20) lines at total pressures ranging from 0.3 to 10 Torr (i.e., CO_2 partial pressures of 27.3–909 mTorr). The signals were fit to the right-hand side of eq 4 multiplied by a term e^{-t} to account for the slow (~ 54 ms^{-1} Torr^{-1}; ref 4) relaxation of 00^01 and diffusion of molecules out of the diode beam. Fitted values

(12) Treanor, C. E., Rich, J. W., Rehm, R. G. *J. Chem. Phys.* 1970, 48, 1966

(13) Dang, C., Reid, J., Garside, B. K. *Appl. Phys.* 1982, B27, 145. Bailly, D., Rossetti, C., Guelachvili, G. *Chem. Phys.* 1985, 100, 101

(14) Bailly, D., Farrenq, R., Rossetti, C. *J. Mol. Spectrosc.* 1978, 70, 124. Bailly, D., Farrenq, R., Guelachvili, G., Rossetti, C. *J. Mol. Spectrosc.* 1981, 90, 74. Bailly, D. 3 eme Cycle Thesis, University of Paris-Sud, 1970
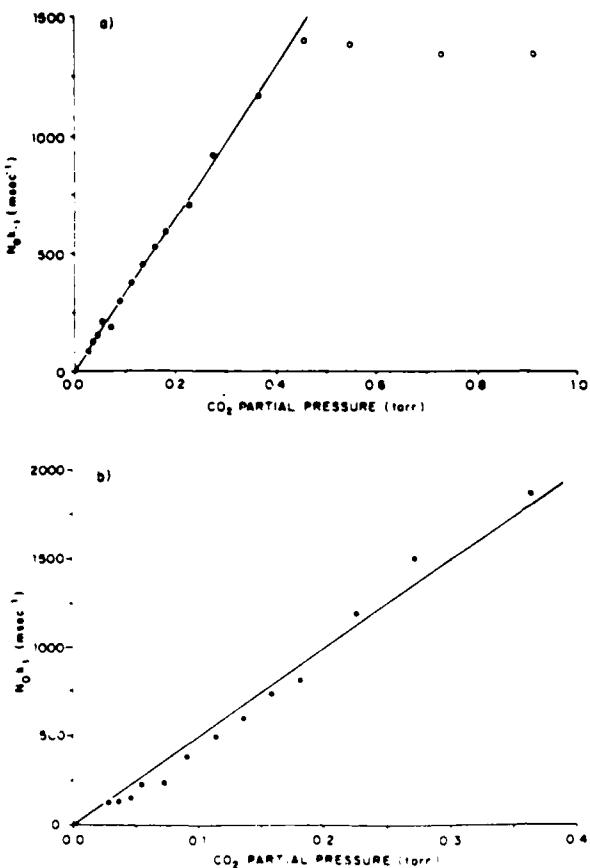
**Figure 3.** Graphs of signal rise rate, $k_{-1}N_0$, versus CO_2 partial pressure for (a) a simple exponential solution $N_2(t)$ given in eq 4 and (b) an induction period solution $N_2^{obsd}(t)$ given in eq 5. Fitted slopes for these graphs are given by (a) $k_{-1} = (3.3 \pm 0.8) \times 10^6$ s^{-1} Torr^{-1} and (b) $k_{-1} = (5.0 \pm 0.6) \times 10^6$ s^{-1} Torr^{-1}. Only solid data points were used in the linear fit of the slope, k_{-1}, and in (b) the intercept was constrained to be equal to zero.

for $k_{-1}N_0$ are graphed as a function of pressure in Figure 3a, and linear behavior is observed up to CO_2 parrtial pressures of 450 mTorr. The slope of these data points is given by $k_{-1} = (3.3 \pm 0.8) \times 10^6$ s^{-1} Torr^{-1}, i.e., a forward rate for eq 1 of $k = (3.7 \pm 0.9) \times 10^6$ s^{-1} Torr^{-1}. The observed rates roll off at a maximum of 1400 ms^{-1}, corresponding to the detector/amplifier response time of ~ 0.7 μs.

Typical (00^02) time domain absorption signals (see for example Figure 1) exhibit a small induction period following the CO_2 laser pulse, suggesting the presence of a secondary process which occurs on the same time scale of the signal of interest. We believe that this feature may be attributed to the finite response (~ 0.7 μs) of the detection system. In order to improve upon the simple data analysis given above, the observed signal, $N_2^{obsd}(t)$, is a modeled by using a standard first-order differential equation to describe the detector/amplifier response, driven by the simple exponential solution for $N_2(t)$ given in eq 4. The observed time domain signal is then given by

$$N_2^{obsd}(t) = N_2(t)\left\{1 - \frac{1}{\delta - k_{-1}N_0}[k_{-1}N_0 e^{-\delta t} - \delta e^{-k_{-1}N_0 t}]\right\} \quad (5)$$

where δ is a free parameter representing the response time of the detection system. The measured (00^02) signals were fitted to the right-hand side of eq 5, multiplied by a term $e^{-\gamma t}$ as before. As shown in Figure 4, $N_2^{obsd}(t)$ exhibits an induction period similar to that seen in the data. Values for $k_{-1}N_0$ obtained from this analysis are graphed as functions of CO_2 partial pressure in Figure
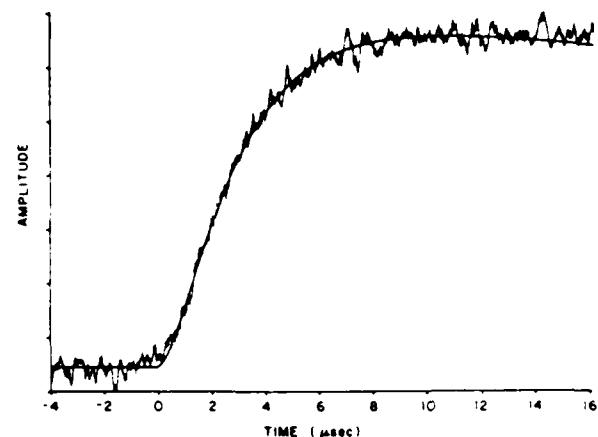
**Figure 4.** Time domain absorption signal for the $(00^02) \rightarrow (00^03)$ R(20) transition for a 1:10 mixture of CO_2 and Ar at a total pressure of 1.0 Torr and a temperature of 298 K. The best fit to this data using eq 5 is shown by the smooth line.

3b. When fitted to a straight line with zero intercept, the fitted slope is $k_{-1} = (5.0 \pm 0.6) \times 10^6$ s^{-1} Torr^{-1}, i.e., a forward rate of $k = (5.6 \pm 0.7) \times 10^6$ s^{-1} Torr^{-1}. The fitted value for δ is fairly constant over a large pressure range, having an average of 1.5 μs^{-1} (a response time of 0.70 μs), exactly the value observed for the roll off of the rates plotted in Figure 3a.

In the fitting model employed above, the (00^01) population is treated as a step function, rising to its maximum value, N_1, immediately upon firing the CO_2 laser. An alternative explanation of the observed induction period may be the presence of significant time dependence in the (00^01) population on the time scale of our measurement. Processes which might be responsible for such a time dependence include (1) incomplete rotational equilibration within the (00^01) level, (2) the finite rise time of the CO_2 laser (measured pulse width ~ 0.4-μs fwhm), and (3) depletion ($\sim 8\%$) of the (00^01) level by the fast process

$$CO_2(00^01) + CO_2(01^10) \rightarrow$$
$$CO_2(01^11) + CO_2(00^00) \quad \Delta\epsilon = -13 \text{ cm}^{-1} \quad (6)$$
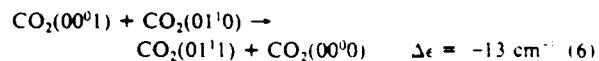
which occurs at a rate of $(5.3 \pm 1) \times 10^6$ s^{-1} Torr^{-1} (ref 9). However, additional modeling of these processes yields essentially the same value for k given above, suggesting that eq 5 is adequate for extracting k from the data.

Discussion

$N_2^{obsd}(t)$ reproduces the induction period seen in the observed time domain (00^02) signals, and as a result, the fitted rise times are faster than those obtained with the simple exponential solution for $N_2(t)$ given in eq 4. The model yields a correct value for the time response of the detection system (as determined in Figure 3a) and provides a small correction to the value for the rate constant obtained via eq 4. Our best estimate for the collisional up-pumping rate, k, is thus $(5.6 \pm 0.7) \times 10^6$ s^{-1} Torr^{-1} or (1.7 $\pm 0.2) \times 10^{-10}$ cm^3 s^{-1} molecule^{-1} at 298 K. This value is quite close to both previous estimates[10,11] and rates for similar near-resonant processes.[7-9] The distorted wave Born calculations of Pack[11] yield a value of $k = 9.2 \times 10^6$ s^{-1} Torr^{-1}, but he suggests that the actual rate is only 0.75 of this value, i.e., 6.9×10^6 s^{-1} Torr^{-1}. In addition, Thomason gives a value of $(7.61 \pm 0.24) \times 10^6$ s^{-1} Torr^{-1} at 700 K, measured by double-resonance using a sequence band CO_2 laser.[10]

In summary, we have measured the rate for collisional up-pumping in CO_2 using the powerful combination of a tunable diode laser spectrometer and CO_2 discharge reference cell. This apparatus provides a high-resolution absorption probe capable of measuring CO_2 populations in rovibrational levels as high as 3 eV in energy. We are presently using this TDL/discharge cell

combination to measure the nascent rotational and vibrational distributions of excited CO_2 molecules created by a variety of chemically interesting processes such as photodissociation,[15] electronic quenching,[16] inelastic scattering,[17] and bimolecular and surface chemical reactions.[18]

(15) O'Neill, J. A.; Kreutz, T. G.; Flynn, G. W., accepted for publication in *J. Chem. Phys.* Wood, C. F.; O'Neill, J. A.; Flynn, G. W. *Chem. Phys. Lett.* **1984**, *109*, 317

(16) Brady, B. B.; Spector, G. B.; Chia, L.; Flynn, G. W. *J. Chem. Phys.* **1987**, *88*, 3245.

(17) O'Neill, J. A.; Wang, C. X.; Cai, J. Y.; Flynn, G. W.; Weston, Jr., R. E. *J. Chem. Phys.* **1986**, *85*, 4195. O'Neill, J. A.; Cai, J. Y.; Flynn, G. W.; Weston, Jr., R. E. *J. Chem. Phys.* **1986**, *84*, 50. Chu, J. O.; Wood, C. F.; Flynn, G. W.; Weston, Jr., R. E. *J. Chem. Phys.* **1984**, *80*, 1703; **1984**, *81*, 5533. Hewitt, S. A.; Hershberger, J. F.; Flynn, G. W.; Weston, Jr., R. E. *J. Chem. Phys.* **1987**, *87*, 1894.

(18) Kreutz, T. G.; O'Neill, J. A.; Flynn, G. W.; Brown, L. S.; Bernasek, S. L., manuscript in preparation.

Direct formation of dielectric thin films on silicon by low energy ion beam bombardment

S S Todorov, C F Yu and E R Fossum, *Department of Electrical Engineering and Columbia Radiation Laboratory, Columbia University, New York, NY 10027, USA*

Dielectric films with thickness of the order of 50 Å are obtained at room temperature by bombarding exposed silicon surfaces with an oxygen-containing ion beam of energy 60 eV. Silicon nitride thin films have also been formed using the same technique. The film thickness is self-limited and largely independent of ion dose. AES and XPS analysis of the produced thin films indicate that the films are not entirely stoichiometric and contain lower oxidation states of silicon. The high electrical quality of the oxide films is demonstrated by the successful fabrication of n-channel MOS transistors with gate dielectrics obtained by ion beam oxidation at room temperature.

Introduction

Silicon dioxide is an indispensable material for silicon MOS technology where its excellent dielectric properties are used both for insulation and field enhancement in the finished devices and at various stages of the fabrication process. Oxides for such applications are usually grown on the surfaces of silicon wafers in oxidizing ambients at temperatures exceeding 900 C. In some applications, such as thin-film FETs on low-melting point substrates, e.g. for flat panel displays, an alternative to thermal oxidation must be found. Reduced-temperature oxidation is also desirable in conventional silicon technology since repeated heating and cooling of the substrate leads to thermal stresses, crystal defects, wafer warpage and impurity redistribution. These become increasingly important with the trend toward smaller device dimensions and the efforts to fabricate stacked three-dimensional MOS structures.

Various techniques have been proposed for oxidation of silicon at reduced temperatures, including plasma oxidation and anodization, chemical vapour deposition (CVD) and plasma-enhanced CVD, reactive sputtering, evaporation or sputtering in an oxygen ambient, and a number of other methods. However, few of these have successfully produced thin oxides suitable for gate dielectrics. Recent reports of successful MOSFET fabrication using oxidation at reduced temperatures include microwave plasma oxidation at 580 C[1], plasma oxidation at 500 C[2] and rf sputter deposition at 200 C for 10 h[3].

In this paper, we report on the use of a broad low-energy ion beam for direct formation of ultra-thin device-quality oxide and nitride films on silicon. The method has allowed the first successful fabrication of FET-quality thin oxides at room temperature. Low energy ion beam oxidation has the added advantage of being a clean vacuum process with independent control over the process parameters and compatibility with other ion beam or vacuum techniques.

Broad low-energy ion beams have previously been applied to the oxidation of metals for fabrication of superconducting[4] and MOM[5] structures and for investigation of reaction kinetics in reactive near-threshold sputtering[6]. Focused ion beams have also been used to study the oxidation kinetics on (111) silicon[7].

Fabrication

The experiments were performed using a single-grid Kaufman-type ion source, shown schematically in Figure 1. The source produces a 2.5 cm dia ion beam and is capable of extracting comparatively large current densities at low applied voltages. The extracted ions impinge on an electrically grounded target positioned 15 cm from the source on a copper substrate holder
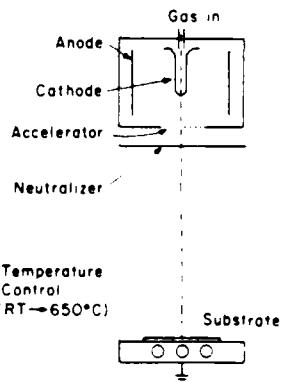
Figure 1. Schematic diagram of low energy ion beam experimental set-up. Base pressure 3 × 10⁻⁷ torr

whose temperature is monitored by a surface mounted thermocouple. The target temperature is observed to rise only several degrees above room temperature during ion beam treatment.

To demonstrate the high electrical quality of the ion beam grown thin dielectrics, MOS transistors were fabricated using the following procedure on 5 cm dia (100) 5 Ω cm p-type silicon wafers, polished on one side. The wafers are chemically cleaned in degreasing agents. H_2SO_4:H_2O_2 and 10% HF. p^+ channel stops and n^- source and drain junctions are formed by high-temperature (1050°C) diffusion of boron and phosphorus, respectively, from spin-on sources through oxide masks prepared by oxidation in steam at 1000°C. A 5000 Å field oxide is grown under the same conditions and gate regions are etched through to the silicon surface using a standard wet buffered oxide etchant (BOE). Immediately before ion beam treatment the wafers are dipped in a cold dilute HF solution to remove any residual native oxide from the gate regions. After ion beam treatment, contact holes are opened through the thin dielectric and 3000 Å of aluminium is evaporated and patterned to form the contact metallization. Then, 3000 Å of aluminium is evaporated on the backside of the wafer to form a substrate contact. Brief post-metallization annealing is carried out for 3 min in N_2 at 400°C.

To investigate the thickness of the thin films. experiments were also performed on simpler substrates on which only MIS capacitors were fabricated according to the above procedure except no p^- and n^+ regions were diffused and no contact holes were etched. These wafers also have large bare regions free of devices used to verify the thickness of the films by ellipsometry measurements.

To ensure that the deposited films are grown by direct ion bombardment and are not due to redeposition of material sputtered from portions of the wafer covered by the field oxide, ion beam oxidation was also carried out on bare silicon wafers cleaned and dipped in HF following the same procedure. The film thickness of these samples was measured on a Gaertner Model L117 ellipsometer immediately after ion beam bombardment and over a period of several days to ensure stability of the deposited films. These samples were also characterized by Auger electron spectroscopy and by XPS in a Leybold-Heraeus LHS-10 surface analysis system equipped with a hemispherical mirror analyser. The base pressure in the characterization chamber is 4×10^{-10} torr. Sputter profiling of the samples is done at 1×10^{-5} torr argon.

Results and discussion

The dependence of the film thickness obtained by ellipsometry on the duration of ion beam treatment is shown in Figure 2. Values for the oxide thickness were also obtained from the accumulation capacitance of MIS capacitors measured after post-metallization annealing. The oxide thickness calculated from $C-V$ measurements before annealing is unreliable due to series resistance in the evaporated aluminium–substrate contacts.

The ellipsometry measurements indicate that after a rapid initial growth rate, the oxide thickness increases very slowly with increasing exposure times. This can be explained by the very limited diffusion of oxygen atoms in the silicon at this low temperature and the inability of fresh oxygen arrivals to reach unreacted silicon. It should be noted that the thicknesses obtained from the $C-V$ measurements are all consistently slightly higher than the values obtained by ellipsometry. This is probably due to amorphization of the thin film by the bombarding ions which
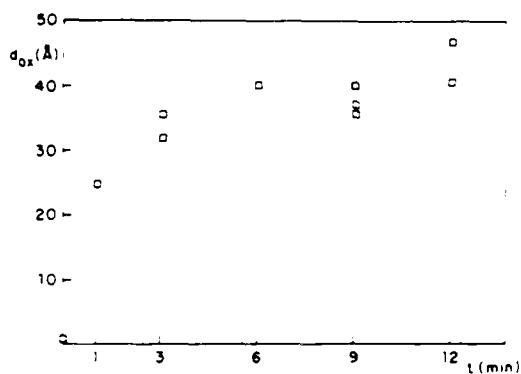
Figure 2. Oxide layer thickness obtained by ellipsometry as a function of ion beam exposure O_2 partial pressure 6.5×10^{-5} torr Ion energy 60 eV.

would lead to the films having a different dielectric constant than that of bulk SiO_2 which is assumed in the calculations of the thickness based on $C-V$ measurements.

This figure is a good illustration of the self-limiting behaviour of the obtained thin films. Self-limitation was observed in the measurements of the resistance of thin films grown by direct ion beam oxidation of nickel[5] and for etch depths obtained by reactive sputter etching[8]. where the self-limitation is believed due to the competing effects of deposition and sputtering.

This dose independence has the useful property of producing very uniform oxide thickness across large targets. The uniformity of the oxides is shown in Figure 3 where it is plotted against the ion beam intensity profile. The oxide thickness is obtained from the accumulation capacitance of 2.8×10^{-3} cm^2 MOS capacitors at various wafer sites. The ion beam intensity is measured by a 1.13 cm dia current probe.

Figure 4 follows the evolution of the Si (LVV) and O (KLL) peaks in the Auger spectrum of an ion beam oxide and a thermal oxide measured by ellipsometry at 47 Å and 48 Å, respectively. The thermal oxide sample was prepared by placing a cleaned silicon wafer in a pure dry oxygen atmosphere at 900 C for 8 min. AES is performed with a 3 keV electron beam having a 1 mm spot size at the target. Profiling is accomplished by sputtering the
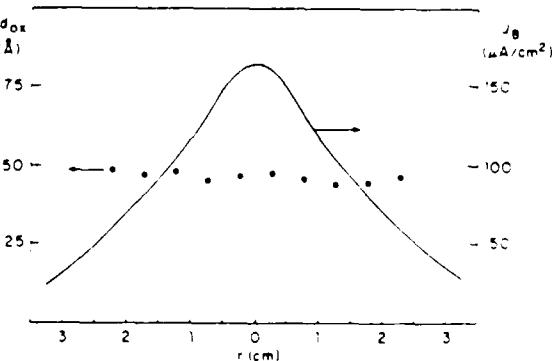
Figure 3. Comparison of oxide thickness uniformity across a 5 cm Si wafer (from $C-V$ measurements of 2.43×10^{-3} cm^2 MIS capacitors) and ion beam current density profile at the target. Current density probe diameter 1.13 cm O_2 partial pressure 6.5×10^{-5} torr Ion energy 60 eV Exposure time 6 min Accelerating voltage 20 V.
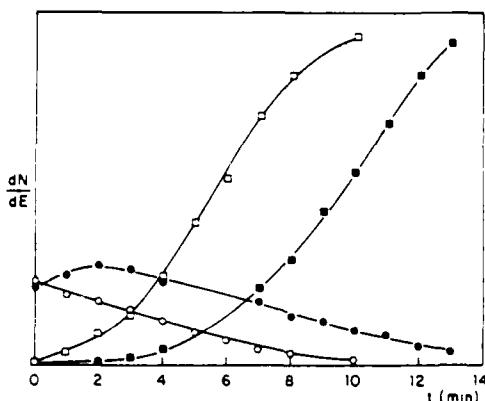
Figure 4. Depth profiles of thermal and ion beam oxides. Plotted are the peak-to-peak amplitudes of Si(LVV) and O(KLL) Auger signals. □—thermal Si; ○—thermal O; ■—ion beam Si; ●—ion beam O

sample with a differentially pumped rastering 3 keV argon ion source with current density 1.2 $\mu A\ cm^{-2}$. Based on ellipsometry, the sputtering rate for SiO_2 is 3.5 Å min^{-1}. The depth profile of the thermal oxide indicates a linearly decaying signal corresponding to uniform composition of the film. The oxygen profile for the ion beam oxide has a peak at a depth of about two monolayers. This, as well as the attenuation of the silicon signal would indicate higher concentration of impurities near the surface. The complete surface Auger spectrum contains a tungsten peak due to sputtering of the source cathode and neutralizer filaments. The shallower slope of the concentration profiles in the ion beam sample compared to the thermal oxide would also indicate a larger width of the interface region probably due to beam induced disorder.

Angle-resolved XPS spectra of a 48 Å thermal and a 47 Å ion beam oxide are shown in Figure 5. XPS is performed using a non-

monochromatized Mg-K_α source providing 1253.6 eV X-rays at a power of 240 W. The analyser is calibrated using the Au 4f7/2 peak with binding energy (BE) 83.8 eV. The calibration and the shifts due to substrate charging are checked throughout the runs by monitoring the C_{1s} peak (BE = 284.5 eV). Both the O_{1s} and Si_{2p} peaks were monitored with progressively shallower sampling depth as the detector was oriented progressively farther away from the surface normal. Figure 5(a) shows the presence of a sharp O_{1s} peak with BE = 533.1 eV at all detector angles which is evidence of the uniform stoichiometry of the thermally grown oxide film. The double Si_{2p} peak includes the silicon signal from the oxide layer (BE = 103.9 eV) and the substrate silicon signal (BE = 99.4 eV). The intensity of the substrate Si_{2p} peak decreases as the sampling depth decreases and disappears at $\theta = 70$. The O_{1s} peak of the ion beam oxide, as shown in Figure 5(b) is much broader and comprises two unresolved components. The higher binding energy one (BE = 532.1 eV) dominates at $\theta = 0$. The lower binding energy peak (BE = 531.6 eV) gradually increases in amplitude relative to the 532.1 eV peak with increasing detector angle and dominates at $\theta = 70$. The behaviour of the lower BE component correlates well with the emergence and growth of a similar lower BE component of the oxide Si_{2p} at higher θ. These lower BE components suggest that the upper surface region consists of oxide with lower oxidation states than SiO_2 due either to the preferential sputtering of the lighter oxygen component or to the stopping depth of the impinging oxygen ions. The ion beam sample Si_{2p} oxide peak has BE = 102.9 eV, a typical value observed for very thin SiO_2 (ref 9).

The high electrical quality of the ion beam grown oxide films is best illustrated by the fabrication of n-channel MOS transistors with gate dielectrics formed by ion beam oxidation at room temperature. The current–voltage characteristics of such a transistor are shown in Figure 6. The gate has a 30 μm length and a 16 μm width. Leakage current through the gate is 6 nA at 1 V, however, this is several orders of magnitude lower than the drain-source saturation current at the voltages at which the transistor
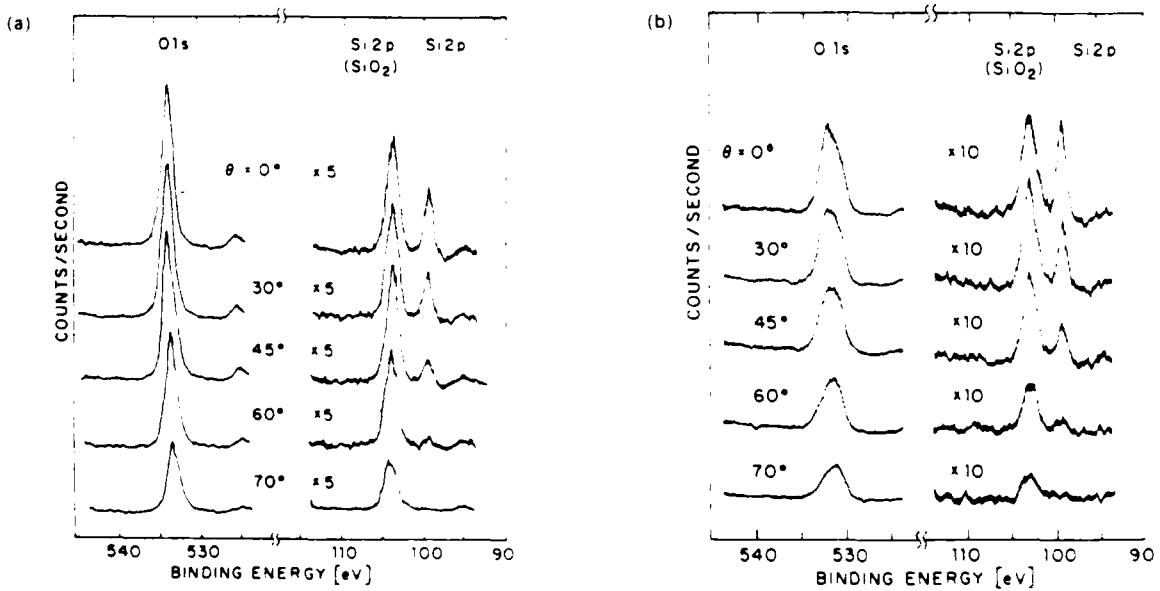

Figure 5. Angle-resolved XPS spectra of (a) thin thermal oxide and (b) thin ion beam oxide. θ is the angle between the surface normal and the detector

931
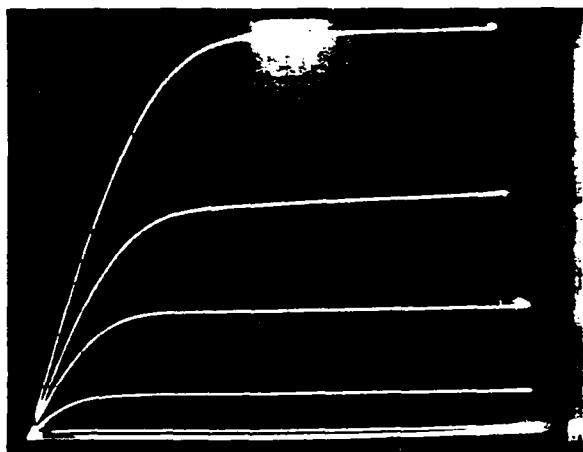
Figure 6. Current–voltage characteristics of a MOSFET fabricated by ion beam oxidation following post-metallization annealing. horiz.—0.2 V-div; vert.—20 μA div. 0.2 V step. O₂ partial pressure 6.5 × 10⁻⁵ torr. Ion energy: 60 eV. Time 12 min.
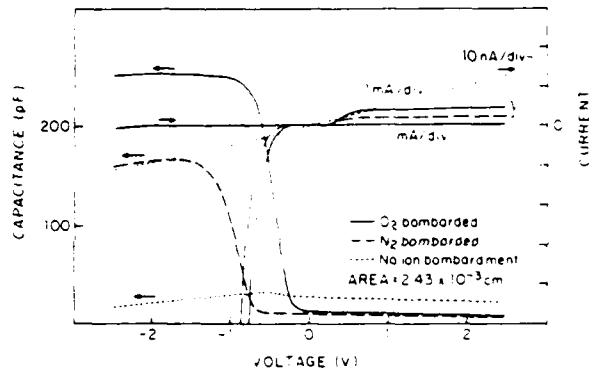

Figure 7. *C–V* and *I–V* characteristics of low energy ion beam oxide (solid curves), nitride (dashed curves), and control (dotted curves) MIS capacitors. 60 eV O₂ for 12 min at 6.5 × 10⁻⁵ torr. 80 eV N₂ for 6 min at 3.2 × 10⁻⁴ torr. Control sample at same conditions as oxide sample but with shutter closed.

operates and does not affect its successful operation. The breakdown strength of the oxide is at least 7×10^6 V cm^{-1}. The saturation mobility is calculated at 575 cm^2 (V s)$^{-1}$ making use of the value of the capacitance in accumulation.

It has been suggested that very thin silicon nitride films may be better dielectrics than silicon dioxide films of comparable thickness. In fact silicon oxynitride films have been successfully produced by 5 keV ion bombardment of silicon[10]. We have demonstrated the ability to produce nitride films on silicon at ion energies below 100 eV. Figure 7 presents a comparison of the *C–V* and *I–V* characteristics of aluminium gate MIS capacitors fabricated by low energy ion beams at room temperature. The control sample was fabricated by the same procedure except that it was not directly bombarded by the ions, since the shutter was not opened. The ohmic nature of the non-bombarded metal–silicon contact is expected and is evidence that not simply charged, but energetic species are necessary for the formation of the thin films. The figure also indicates that reverse-bias leakage current of the nitride sample is lower than that of the oxide by a factor of two.

Summary

Device-quality ultra-thin dielectric films have been produced on silicon substrates bombarded by reactive ions with energies 60–80 eV. The damage induced by the ion beam does not significantly affect device operation.

Acknowledgement

References

[1] S-I Kimura, E Murakami, T Warabisako, H Sunami and T Tokuyama, *IEEE Electron Device Letts*, **EDL-7**, 38 (1986)

[2] A K Ray and A Reisman, *J Electrochem Soc*, **128**, 2424 (1981)

[3] H-S Lee and S-C Chang, *IEEE Electron Device Letts*, **EDL-3**, 310 (1982)

[4] A W Kleinsasser and R A Buhrman, *Appl Phys Letts*, **37**, 841 (1980)

[5] J M E Harper, M Heiblum, J L Speidell and J J Cuomo, *J appl Phys*, **52**, 4118 (1981)

[6] T M Mayer, J M E Harper and J J Cuomo, *J Vac Sci Technol*, **A3**, 1779 (1985)

[7] H Daimon and Y Murata, *Japan J appl Phys*, **21**, L718 (1982)

[8] S Berg, B Gelin, M Ostling and S M Babulanam, *J Vac Sci Technol*, **A2**, 470 (1984)

[9] P J Grunthaner, M H Hecht, F J Grunthaner and N M Johnson, *J appl Phys*, submitted for publication (1986)

[10] W Streb and R Hezel, *J Vac Sci Technol*, **B2**, 626 (1984)

Reduced reverse bias current in Al–GaAs and $In_{0.75}Ga_{0.25}As$–GaAs junctions containing an interfacial arsenic layer

D. V. Rossi and E. R. Fossum

Microelectronics Sciences Laboratories, 1312 S. W. Mudd Building, Columbia University, New York, New York 10027

G. D. Pettit, P. D. Kirchner, and J. M. Woodall

IBM T. J. Watson Research Center, Yorktown Heights, New York 10598

Interfacial As is shown to reduce reverse-bias current in Al–GaAs Schottky barriers. It is suggested that the leakage reduction is associated with the removal of low work function phases at the interface. In addition, current–voltage measurements performed on $In_{0.75}Ga_{0.25}As$–GaAs heterojunctions indicate a dependence upon the condition of the GaAs prior to deposition of the $In_{0.75}Ga_{0.25}As$ layer.

I. INTRODUCTION

Excess reverse bias current in metal–semiconductor contacts can degrade device performance. For example, gate current limits field-effect transistor (FET) sensitivity and increases power consumption. In charge-coupled device applications, gate current can contribute to dark current and introduce nonlinearity in charge packets, thus reducing the signal-to-noise ratio. Reverse bias current reduction in GaAs devices requires insight into the relationship among chemical, structural, and electrical properties of the metal-semiconductor interface. A manifestation of the interaction among these properties is the phenomenon of Fermi-level pinning.

Understanding Fermi-level pinning, however, remains a formidable problem. Models which attempt to describe this behavior include metal-induced gap states (MIGS),[1] extrinsic defect generation during adatom deposition,[2] and effective work functions of mixed interface phases resulting from chemical reactions between metal and semiconductor.[3] The effective work function (EWF) model asserts that these reactions primarily result in excess anion precipitates (e.g., arsenic) which dominate the interface behavior and thus dictate the effective work function. If low work function regions exist at the unannealed interface they could reduce the effective work function and observed barrier height, noticeably increasing the reverse bias current. Ohmic behavior, for example, has been reported from a low-work function phase of Au–Ga that forms at the periphery of an annealed Au–GaAs contact.[4] Reduction or elimination of these low-work function areas should reduce the reverse bias current. This paper reports an attempt to reduce reverse bias current in Al–GaAs Schottky diodes by creating a homogeneous interface through deposition of an arsenic layer prior to metallization. In addition, the electrical properties of an $In_{0.75}Ga_{0.25}As$–GaAs junction have been examined and are shown to be dependent upon the condition of the GaAs surface prior to deposition of the InGaAs layer.

II. Al–GaAs JUNCTION

In this experiment, a molecular beam epitaxy (MBE) system is used to grow an Al–As–GaAs "test" structure and an Al–GaAs "control" structure. After initial preparation, an n^+ GaAs sample was mounted onto a Mo block using In and loaded into the MBE chamber. Subsequent heating removes the native oxide and helps form a backside ohmic contact. A 0.5-μm, 2×10^{18} cm^{-3} n-GaAs buffer layer is grown followed by an additional 0.5-μm n-GaAs layer doped at 2×10^{17} cm^{-3}. The "test" structure was allowed to cool to room temperature and then exposed to an As$_2$ flux for a time corresponding to the deposition of 100 Å of As, assuming a unity sticking coefficient. Prior to removal from the chamber, 1800 Å of Al was evaporated on both samples. Diodes were defined by photolithography and wet chemical etching. The maximum temperature during these latter procedures was 90 °C which should avoid any annealing effects that could degrade the As interlayer's homogeneity.

Current–voltage (I–V) characteristics were measured at room temperature, and the barrier height and ideality calculated from the forward bias data using a least-squares fit. A typical comparison between the "test" and "control" structures is depicted in Fig. 1, indicating that the sample contain-
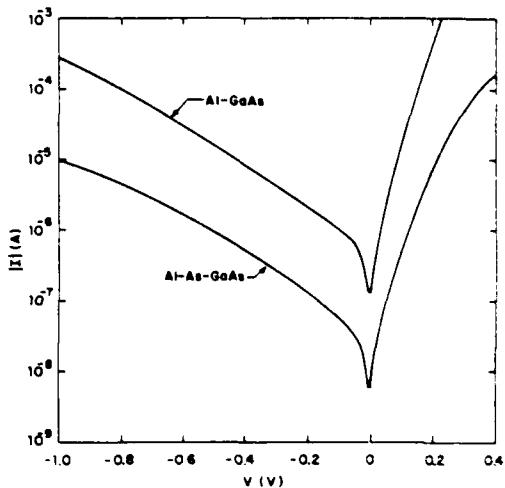

FIG. 1. Comparison between Al–GaAs and Al–As–GaAs I–V characteristics showing large reduction in reverse leakage current.

TABLE I. Ideality factor, barrier height, and leakage current density at 1 V reverse bias for representative samples.

Sample	Barrier height (eV)	Ideality	Current density at −1 V (mA/cm²)
Al/GaAs	0.65	1.06	31.12
Al/As/GaAs	0.70	1.20	1.10
1(a)	0.65	1.47	0.86
1(b)	0.60	1.03	3.53
2(a)	0.66	1.01	0.11
2(b)	0.51	1.07	70.70

ing the As interlayer exhibits a 25-fold reduction of reverse-bias current. Table I compares the barrier height, ideality, and current density for the two samples. The test structure, while displaying less ideal behavior, does in fact have a typically 50–60 mV greater barrier height to accompany the reduced current. The difference in barrier heights has been confirmed by internal photoemission measurements. Although roughly an 80 mV barrier height enhancement would be necessary to account for the current reduction, the measured barrier height difference is within the experimental error. Reverse bias current reduction, therefore, has been achieved by generating a more uniformly pinned interface.

Indeed, the coincidence of As and reduced-dark current has been previously observed[5]. Ostensibly contradictory data for Al on arsenic-rich GaAs surfaces also exists.[6] Wang showed that increasing As-surface coverage is associated with a concomitant decrease in the Al-barrier height to GaAs. It was also shown, however, that the opposite is true for Al on AlAs possibly due to the fact that intermixing of Al and Ga is precluded in the AlAs case. It is likely that Al and Ga intermixing does not occur for the thick-As layer used in this experiment and thus, the apparent contradiction can be explained.

It might also be possible to associate the reduced current with a metal–insulator semiconductor (MIS) model of the junction. This effect is improbable since the current reduction is a constant factor nearly independent of the applied voltage. Experiments to differentiate between various models of the reverse current transport mechanism are in progress.

III. $In_{0.75}Ga_{0.25}As$–GaAs HETEROJUNCTION

The influence of the GaAs-surface condition on the measured InGaAs–GaAs barrier height was also investigated. MBE was used to grow a 0.5-μm-GaAs buffer layer on a semi-insulating substrate followed by a 0.5-μm, 5×10^{16} cm^{-3} n-GaAs layer. At this point, four different sample preparations followed:

1(a) Approximately 100 Å of As was deposited at room temperature between the GaAs and the intrinsic InGaAs, the latter of which was grown at 200 °C.

1(b) Intrinsic InGaAs was grown on the GaAs at 450 °C.

2(a) The sample was removed from the chamber and air exposed. It was then reinserted and intrinsic InGaAs grown at 200 °C.

2(b) This sample was also air exposed. The InGaAs-growth temperature, however, was 350 °C which should desorb the As oxide at the GaAs surface.

In all samples, the intrinsic InGaAs layer is 0.2-μm. The samples are schematically represented in Fig. 2. After the InGaAs growth, 180-μm-diam Au contacts were evaporated onto the InGaAs through a stainless steel mask in a deposition system having a base pressure of 1×10^{-7} Torr. Probing between Au contacts before mesa etching the InGaAs layer yields ohmic behavior indicative of a low barrier contact of the Au to the intrinsic InGaAs. After mesa etching, the I–V characteristic resembles that of two back-to-back diodes. Ohmic contacts were formed by alloying In to the GaAs at 300 °C in forming gas using a rapid thermal annealer. Before measuring the rectifying behavior of the

InGaAs growth Temperature: 200°C

Sample 1a

InGaAs growth Temperature: 200°C

Sample 2a

InGaAs growth Temperature: 450°C

Sample 1b

InGaAs growth Temperature: 350°C
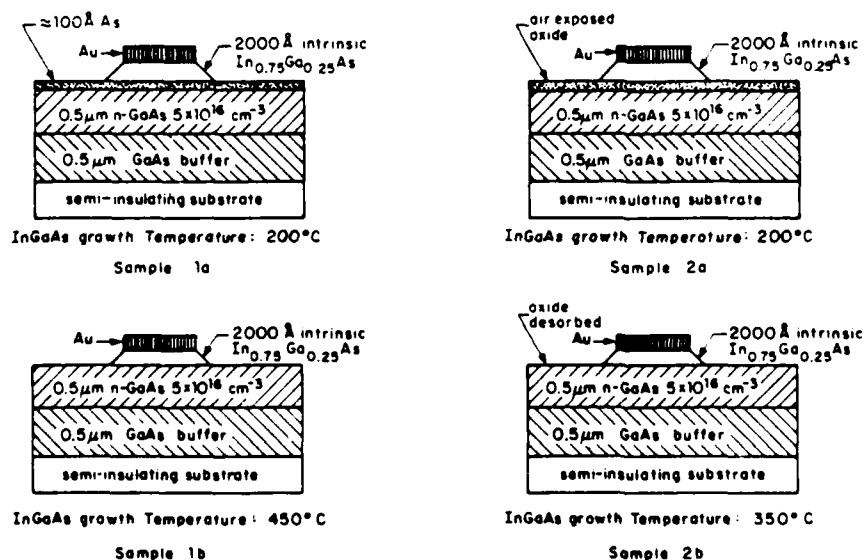
Sample 2b

FIG. 2. Schematic cross section of the $In_{0.75}Ga_{0.25}As$–GaAs junctions. Samples 1(a) and 1(b) were grown entirely under UHV conditions. Samples 2(a) and 2(b) were air-exposed prior to InGaAs deposition. Au was evaporated after removal from the MBE chamber.
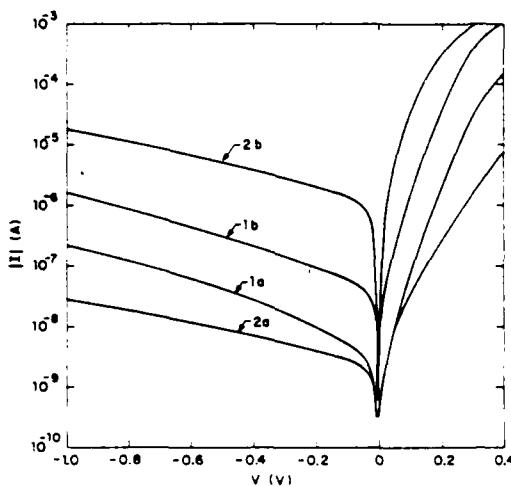
FIG. 3. *I–V* characteristics for $In_{0.75}Ga_{0.25}As$–GaAs junctions indicating a strong dependence on the GaAs surface condition.

InGaAs–GaAs junction, the characteristic between two Au contacts as well as the integrity of the In contacts were confirmed.

Typical *I–V* characteristics for the four samples are shown in Fig. 3, while the barrier heights and ideality factors are summarized in Table I. The samples possessing interfacial As, whether elemental or in oxide form, exhibit current reduction. Specifically, the InGaAs–As–GaAs sample has four times less current at 1-V-reverse bias than the InGaAs–GaAs case. More dramatically, for the air-exposed samples the one in which the As oxide was desorbed has roughly a factor of 640 greater current than the sample grown at lower temperature. In fact, the air-exposed low temperature growth sample has the lowest reverse bias current of all samples tested as well as the best ideality factor. Furthermore, the barrier heights for the InGaAs–As–GaAs

sample and the air-exposed sample possessing the As oxide are nearly equivalent to within the 10-mV deviation in measurement. The lowest barrier height is associated with the air-exposed sample in which the As oxide was desorbed.

IV. SUMMARY

Leakage current in Al–GaAs Schottky diodes has been reduced by the introduction of an interfacial As layer. While this effect does not preclude other models of the metal–semiconductor interface, it does support the premises of the effective work function model in which the barrier height reflects an average work function of mixed phases and As is believed to pin the interface Fermi level. The measured barrier height of the $In_{0.75}Ga_{0.25}As$–GaAs heterojunction has also been shown to depend upon interfacial As, regardless of the As phase. Investigation of leakage current reduction not only lends insight into the interface but may also have practical implications for threshold voltage control and sidegating in GaAs integrated circuits.

[1] J. Tersoff, Phys. Rev. B **32**, 6968 (1985).
[2] W. E. Spicer, I. Lindau, P. Skeath, and C.Y. Su, J. Vac. Sci. Technol. B **17**, 1019 (1980).
[3] J. L. Freeouf and J.M. Woodall, Appl. Phys. Lett. **39**, 727 (1981).
[4] Z. Liliental-Weber, R. Gronsky, J. Washburn, N. Newman, W. E. Spicer, and E. R. Weber, J. Vac. Sci. Technol. B **4**, 912 (1986)
[5] H. J. Stocker and D. E. Aspnes, Appl. Phys. Lett. **42**, 85 (1983).
[6] W. I. Wang, J. Vac. Sci. Technol. B **1**, 574 (1983)
[7] P. Oelhafen, J. L. Freeouf, G. D. Pettit, and J. M. Woodall, J. Vac. Sci. Technol. B **1**, 787 (1983).

to be published JVST

GROWTH MECHANISM OF THIN OXIDE FILMS UNDER LOW ENERGY OXYGEN ION BOMBARDMENT

S.S. Todorov* and E.R. Fossum
Department of Electrical Engineering and
Columbia Radiation Laboratory
Columbia University
New York, NY 10027

* Permanent Address: Institute of Electronics, Sofia, Bulgaria

Abstract

Bombardment of silicon surfaces by low energy oxygen ions has been investigated as a possible process for growing films of SiO_2 at room temperature. Broad ion beams of energy 40-200 eV and variable oxygen content have been used to grow ultra-thin oxides of extremely uniform thickness. The ion beam oxides are similar to thin thermal oxides in many respects - composition, chemical binding, optical and electrical properties. The dependence of the thickness and quality of the oxide films on ion dose, ion energy and substrate temperature have been investigated. The obtained thickness is observed to vary only slightly with increasing substrate temperature up to 650°C which indicates non-thermal process kinetics. The ion-beam oxides reach a limiting thickness of 40-60 Å which is largely independent of ion dose and is also found to be insensitive to ion energy.

The observed oxidation rates are explained on the basis of radiation-enhanced diffusion and reaction processes. Limited thicknesses are observed even when sputtering is negligible because of the decreasing effective penetration of the ions due to the swelling of the target which accompanies the conversion of Si to SiO_2. Thus the film grows until the oxide-semiconductor interface moves beyond the current ion penetration depth after which oxidation effectively stops. This model is equally applicable to high-energy high-dose oxygen ion implantation for production of buried oxides in SOI technology where it is observed that oxide growth occurs predominantly at the upper interface.

1

I. INTRODUCTION

Further advances in the development of high density integrated circuits (ICs) will require scaled down device dimensions, closer device packing and a greater number of fabrication steps. These in turn require reduced IC fabrication thermal budgets and, in the case of metal-oxide-semiconductor (MOS) devices, thinner gate dielectrics. The trend toward larger substrates places an additional requirement on the lateral uniformity of the gate oxides. The problem of obtaining high quality thin films of SiO_2 at reduced temperatures has been approached from two directions - oxide growth and oxide deposition. These differ in the type of material being supplied to the substrate - oxygen only or both oxygen and silicon. Both approaches rely on the use of some form of energetic species to enhance the low oxidation rates typically observed at reduced temperatures. Successful fabrication of MOS devices has been reported for gate dielectrics grown by plasma oxidation at 500-600°C [1,2] and deposited by rf sputtering at 200-300°C [3]. Very thin SiO_2 films of high electrical quality have also been deposited by plasma enhanced chemical vapor deposition (PE-CVD) at 350°C [4]. MOS transistors have been fabricated using very thin gate dielectrics grown at 25°C by ion beam oxidation [5].

Due to the spatial separation of the plasma producing region and the sample, ion beam oxidation permits independent control over ion energy and ion flux, substrate temperature and angle of incidence. These in turn permit detailed study of the influence

2

of the different process parameters on the obtained oxides, the results of which may be extended to rf oxidation.

The electrical characteristics and other properties of the very thin oxide films grown at room temperature by ion beam oxidation have been reported previously [5,6]. Briefly, these showed that ion beam oxides are similar to thermally grown SiO_2 films of equal thickness in many respects. They have almost identical Auger sputter profiles except for the apparently sharper interface of the thermal oxide. X-ray photoelectron spectroscopy at glancing detector angles indicated the presence of lower oxides of silicon at the film surface. The most notable difference between thermal and ion beam oxides is the 2-3 orders of magnitude greater leakage current in reverse bias for the latter. However, this leakage was another three orders of magnitude lower than the drain-source saturation current and thus did not degrade MOSFET performance [5]. In this paper we will discuss the dependence of the oxide growth kinetics on the ion bombardment conditions and will propose a mechanism describing the oxidation process.

II. EXPERIMENTAL

The test vehicles are (100) silicon wafers, 5-cm diameter, polished on one side. Both p-type and n-type wafers of resistivity 2-5 $\Omega \cdot cm$ are used. The samples are prepared by chemical cleaning and the growth and stripping of two thick $900^{\circ}C$ oxides after which they are loaded into the ion beam processing chamber. Ion beam oxidation is performed using a

single-grid broad-beam ion source, described previously [7]. The
source operates on a mixture of argon and oxygen. It was found
that a 1:1 Ar:O_2 ratio allowed both good control over source
operation for extended periods of time and led to the production
of oxide films on the substrates. After unloading from the ion
beam processing chamber the thickness of the obtained films is
measured with a Gaertner L117 ellipsometer using a He-Ne laser.
The thicknesses measured by ellipsometry correlate well with
those obtained by Auger sputter profiling, calculated from the
ratio of the XPS Si 2p peaks or calculated from the capacitance-
voltage characteristics of MOS capacitors [6]. The oxide
thickness was calculated from the ellipsometer measurements with
the standard a priori assumption of the refractive index
necessary for ultra thin films. The index of refraction used is
that of thermal SiO_2. This is justified by the aforementioned
similarity between the two types of oxides, further illustrated
in Fig. 1 which compares the unprocessed ellipsometer parameters
Δ and ψ . The solid line shows the evolution of Δ and ψ for
thermal oxides of increasing thickness grown at 875°C. The
numbers represent ion beam oxides and correspond to the oxygen
ion dose received during the exposure. The ion beam current
density was 135 $\mu A/cm^2$ which corresponds to a flux rate of
5×10^{16} $O/cm^2 \cdot min$. It is clear that the two oxides evolve
along essentially the same lines pointing further to their
similarity.

4

III. RESULTS AND DISCUSSION

The dependence of the obtained ion beam oxide thickness on oxygen ion dose is presented in Fig. 2. The silicon wafers were bombarded by a 60 eV ion beam containing O_2 and Ar in a 1:1 ratio. The substrate temperature was monitored by a substrate-holder mounted thermocouple and is 25°C. The figure shows that there is an approximate logarithmic dependence on the oxygen ion dose though the scatter in the data is more pronounced at higher doses. This scatter is attributed to instabilities in the ion source which become more important at the long exposure times necessary to achieve higher doses. It appears that the n-type wafers oxidize somewhat more slowly than the p-type wafers, however, the differences are within the experimental uncertainty. It has been observed previously that the ion beam oxides exhibit very good thickness uniformity over the surface of the wafer even though they are grown by an ion beam which has a peaked gaussian-like profile [5]. This insensitivity to oxygen ion dose was attributed to the attaining of a self-limiting value of the thickness. While it is now clear that such self-limiting behavior is not being observed, it is also clear that the growth rate at high doses is sufficiently low so that even two-fold increases in the ion dose would lead to the growth of only about a monolayer of oxide.

Self-limiting oxide growth has been observed in the case of rf oxidation of lead [8] and ion beam oxidation of nickel [9]. It is attributed to the achieved balance between the competing processes of oxidation and sputtering. The oxidation rate is

5

initially large and decreases with increasing film thickness whereas the sputtering rate is independent of oxide thickness. Thus the oxide growth will stop at a thickness value at which the two rates become equal.

It is expected that self-limitation will occur earlier at higher ion beam energies where sputtering is more pronounced. That is indeed the case as shown in Fig. 3. The samples are bombarded with a 100 eV ion beam with a 1:1 O_2:Ar ratio at 25°C. It is seen that at higher ion energies higher initial oxidation rates are observed. This is probably related to the greater depths at which the ions are stopped and the increased probability that they will be retained in the target. The oxidation rate then decreases more quickly than for 60 eV ions and appears to exhibit self-limiting behavior.

To investigate the importance of beam heating of the substrate during ion bombardment several samples were oxidized at different oxygen dose rates leading to the same received dose. The ion flux at the substrate was varied by changing the source discharge power or by changing the source-to substrate distance. No differences in the properties of the obtained oxides were observed.

The substrate may be intentionally heated up to 650°C by means of resistive heating of the substrate holder. The obtained oxide thickness for fixed ion beam conditions is observed to increase with increasing substrate temperature. However, this dependence is rather weak, especially considering the large temperature range. Presenting the data in an Arrhenius plot, as

6

shown in Fig. 4, allows the extraction of an activation energy E_{act} for the process. Here the oxidation rate dependence has been approximated by $x = C \ln t$, where t is the oxidation time and C is a constant which is assumed to depend on the temperature T in the standard fashion: $C = C_0 \exp(-E_{act}/kT)$, where k is the Boltzmann constant. Note that the data is largely independent of ion dose. The slope of the obtained straight line allows the determination of $E_{act} = 7$ meV. For comparison the activation energy of the linear rate constant B/A defined in thermal oxidation of silicon is 2 eV [10]. Clearly, oxidation due to ion bombardment relies on non-thermal mechanisms of enhancing the diffusivity and reactivity of the oxidizing species.

The observed oxide thicknesses correlate well with the expected oxygen ion projected range, accounting for the volume expansion which accompanies the conversion of silicon to SiO_2. Thus, the obtained thickness may be expected to increase at higher ion energies. The dependence of oxide thickness on ion energy for a fixed oxygen ion dose is shown in Fig. 5 for beams with a 50% oxygen content. The lack of a trend over the covered energy range is evident. It is difficult to imagine that the increased oxidation rate is exactly counteracted by an equal increase in the sputtering yield unless both processes are governed by the same mechanism.

7

IV. PROCESS MODEL

Previous models of low energy ion beam oxidation of metals describe the oxidation process in the absence of sputtering as decreasing exponentially with increasing exposure time [9]. However, this model is carried over from the process of rf oxidation of metals in which a self-biasing of the oxide layer is observed [8,11] and does not account for the fact that the oxidizing species are actually deposited below the surface and are distributed throughout the film. Also it does not explain the observed enhancement in the oxidation rates over thermal oxidation processes.

It is considered appropriate to view low-energy ion beam oxidation of silicon as a scaled-down version of the high-energy high-dose oxygen implantation processes used for obtaining buried layers of SiO_2 in silicon-on-insulator (SOI) technology. The models developed to describe this process introduce enhanced effective diffusion constant and reaction rate and also take into account the swelling of the target [12,13]. These models, however, neglect any spatial dependence of the enhancement and lead to the prediction of preferential growth toward the substrate [12]. It has been observed that the growth of the buried oxide layer takes place mostly at the top interface [14,15]. The preferred oxidation upwards has been attributed to the different fate of silicon interstitials at the two $Si-SiO_2$ interfaces [16]. The interstitials which are emitted during the oxidation process can recombine at defects or at the surfaces of the wafer. Due to the much larger distance to

8

the back surface, the interstitials at the bottom interface are practically fixed and serve to block the oxidation process [16].

We propose an important modification to these models. While it is clear that sputtering and oxidation compete for domination in these processes, the explanation of enhanced oxidation in unheated substrates is considered unsatisfactory. It is proposed that the enhanced reactivity and mobility of the oxidizing species are due to the ion bombardment. Thus, at any time, oxidation is taking place only at those depths which are currently being reached by the ion beam. The swelling which accompanies the conversion of silicon to silicon dioxide effectively prevents later ion arrivals from reaching the same depths in the silicon. Moreover, it is difficult for them to diffuse deeper into the substrate after they are stopped since their diffusion is enhanced only in the region in which the ions are depositing energy. In other words, it is important to recognize that the effective process rates are spatially dependent. Because of the volume expansion this spatial dependence means that the oxidation rate at the oxide-semiconductor interface is time dependent and oxidation at the lower interface will stop when a sufficient portion of the substrate has been converted to SiO_2. Note that the concept of a temperature increase due to thermal spikes or beam heating is not necessary for this treatment. Enhanced mobility and reactivity may be due to other beam related processes like the creation of excess numbers of defects in the vicinity of an atom or the transfer of excess energy to the atom in a collision [17].

The proposed model accounts for the enhanced oxidation rates observed for unheated substrates under ion bombardment. Also, it accounts for the insensitivity to substrate temperature observed in experiments on low energy ion beam oxidation of silicon. Further, it naturally explains the limited thicknesses observed with this technique even in the absence of sputtering. The model is simply extended to the preferred growth toward the top surface of the wafer observed in the case of buried oxides produced by high energy oxygen implantation. A quantitative evaluation of the process parameters based on this model will be published separately [18].

V. CONCLUSION

In summary the bombardment of silicon surfaces by low energy oxygen ion beams is observed to lead to the growth of ultra thin films of silicon dioxide. The oxide growth is self-limiting due to the competing processes of oxidation and sputtering. Limited thicknesses are also observed in the absence of sputtering. The oxidation process is explained in terms of beam-enhanced reactivity and mobility of the oxidizing species. The importance of recognizing the spatial limitations of the enhanced oxidation process is stressed. The explanation of the process is extended to buried SiO_2 layers produced by oxygen implantation.

VI. ACKNOWLEDGEMENT

10

REFERENCES

1   A.K. Ray and A. Reisman, J. Electrochem. Soc., 128, 2424 (1981).

2   S.I. Kimura, E. Murakami, T. Warabisako, H. Sunami, and T. Tokuyama, IEEE Electron Device Lett., EDL-7, 38 (1986).

3   H.S. Lee and S.C. Chang, IEEE Electron Device Lett., EDL-3, 310 (1982).

4   J. Batey, E. Tierney, and T.N. Nguyen, IEEE Electron Device Lett., EDL-8, 148 (1987).

5   S.S. Todorov, S.L. Shillinger, and E.R. Fossum, IEEE Electron Device Lett., EDL-7, 468 (1986).

6   C.F. Yu, S.S. Todorov, and E.R. Fossum, Proc. IVC-10,ICSS-6, 33rd AVS, Baltimore, Md, Oct. 1986. (to be published in J. Vac. Sci. Technol.)

7   S.S. Todorov, C.F. Yu, and E.R. Fossum, Vacuum, 36, 929 (1986).

8   J.H. Greiner, J. Appl. Phys., 42, 5151 (1971).

9   J.M.E. Harper, M. Heiblum, J.L. Speidell, and J.J. Cuomo, J. Appl. Phys., 52, 4118 (1981).

10   B.E. Deal, J. Electrochem. Soc., 110, 122 (1979).

11   A.T. Fromhold and M. Baker, J. Appl. Phys., 51, 6377 (1980).

12   H.U. Jager, E. Hensel, U. Kreissig, W. Skorupa, and E. Sobeslavsky, Thin Solid Films, 123, 159 (1985).

13   E.A. Maydell-Ondrusz and I.H. Wilson, Thin Solid Films, 114, 357 (1984).

14 T. Hayashi, H. Okamoto, and Y. Homma, Jpn. J. Appl. Phys.,
    19, 1005 (1980); T. Hayashi, S. Maeyama, and S. Yoshii, Jpn.
    J. Appl. Phys., 19, 1111 (1980).

15 J.A. Kilner, R.J. Chater, P.L.F. Hemment, R.F. Peart, E.A.
    Maydell-Ondrusz, M.R. Taylor, and R.P. Arrowsmith, Nucl.
    Instr. Meth., B7/8, 293 (1985).

16 A.H. van Ommen, B.H. Koek, and M.P.A. Viegers, Appl. Phys.
    Lett., 49, 628 (1986).

17 S.M. Myers, Nucl. Instr. Meth., 168, 265 (1980).

18 S.S. Todorov and E.R. Fossum, to be published.

FIGURE CAPTIONS

Fig. 1. Comparison of the evolution of thin 875°C thermal (solid line) and 25°C ion beam (numbers) oxide films. The numbers correspond to the exposure in minutes to a 60 eV ion beam of 1:1 $Ar:O_2$ content and an oxygen dose rate of 5×10^{16} $cm^{-2}min^{-1}$.

Fig. 2. Ion beam oxide thickness as a function of oxygen dose. Ion energy: 60 eV. Beam oxygen content: 50 %. Substrate temperature: 25°C. □ - p-type silicon; ▨ - n-type silicon.

Fig. 3. Ion beam oxide thickness as a function of oxygen dose. Ion energy: 100 eV. Beam oxygen content: 50 %. Substrate temperature: 25°C.

Fig. 4. Arrhenius plot for logarithmic ion beam oxidation rate constant. Ion energy: 60 eV. Beam oxygen content: 50 %. □ - O dose: 1.3×10^{17} cm^{-2}. ◇ - O dose: 2.7×10^{17} cm^{-2}.

Fig. 5. Ion beam oxide thickness as a function of ion energy. Beam oxygen content: 50 %. Substrate temperature: 25°C. □ - O dose: 3.0×10^{17} cm^{-2}. ◇ - O dose: 5.3×10^{17} cm^{-2}.

13

Oxide thickness (A) vs Ion beam energy (eV)
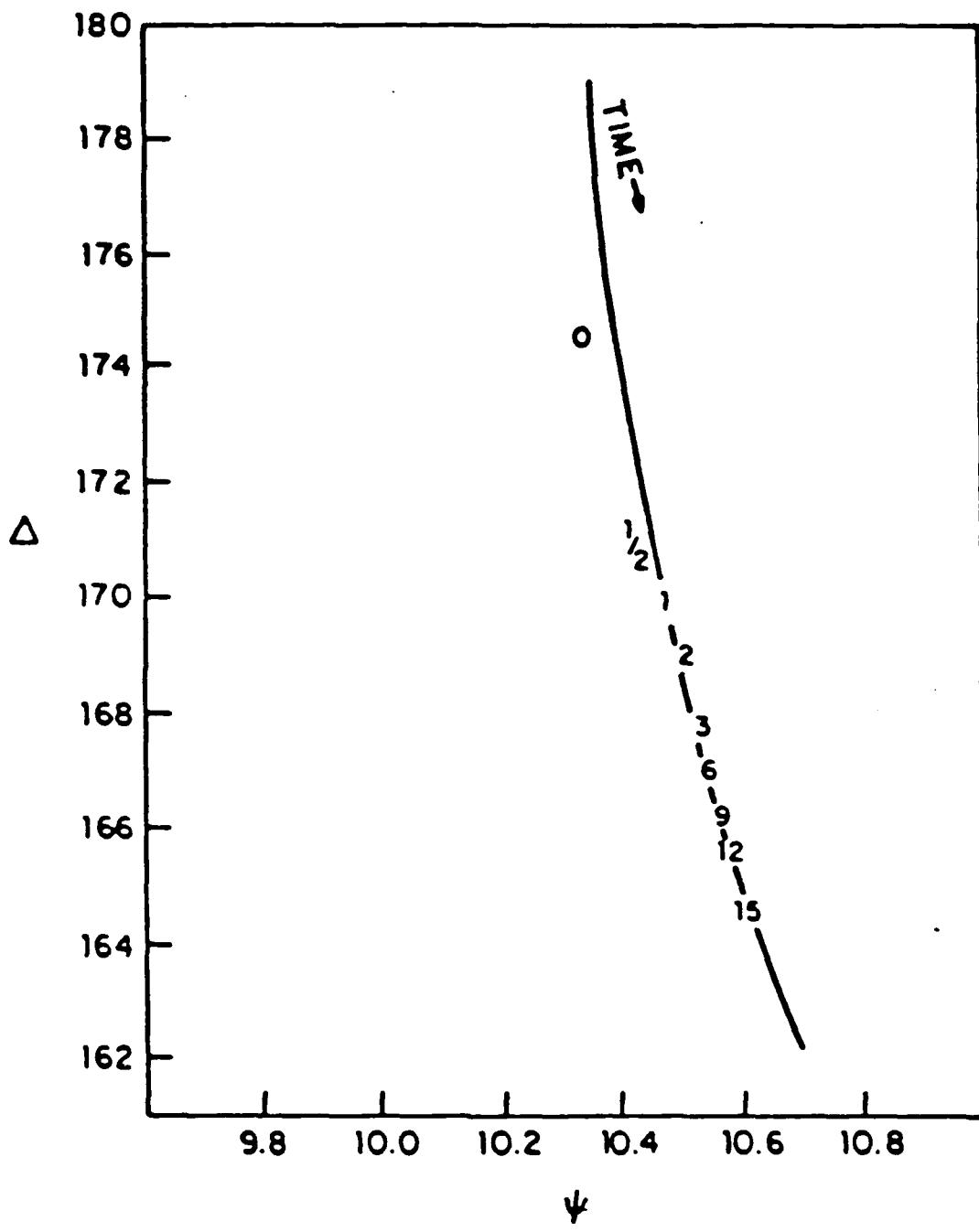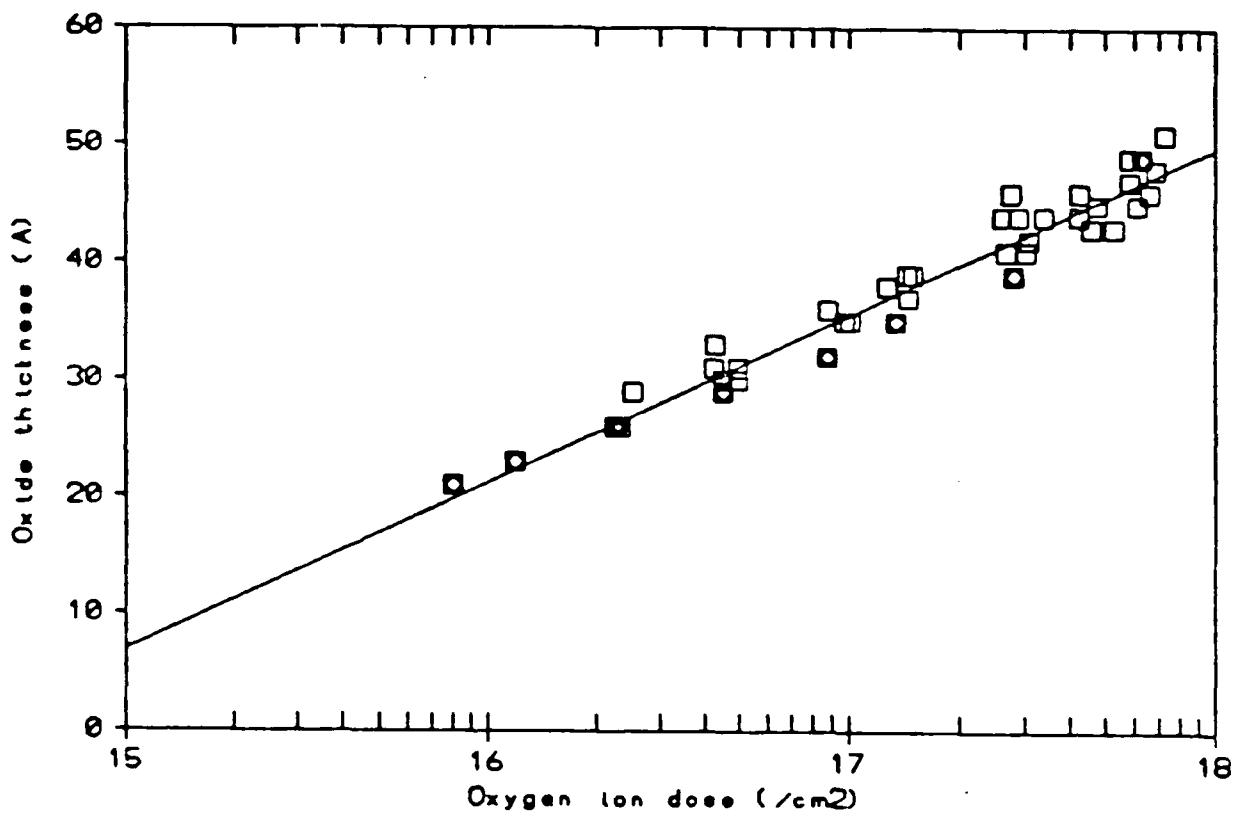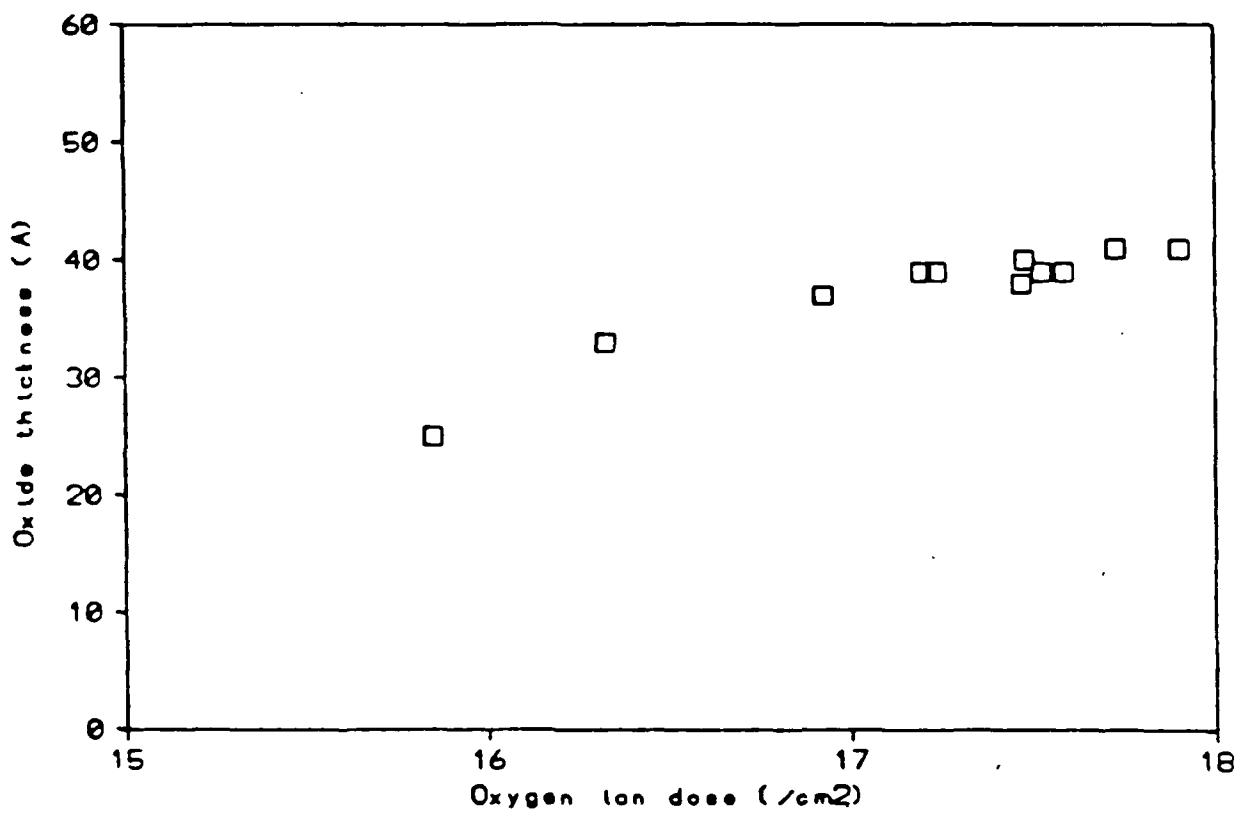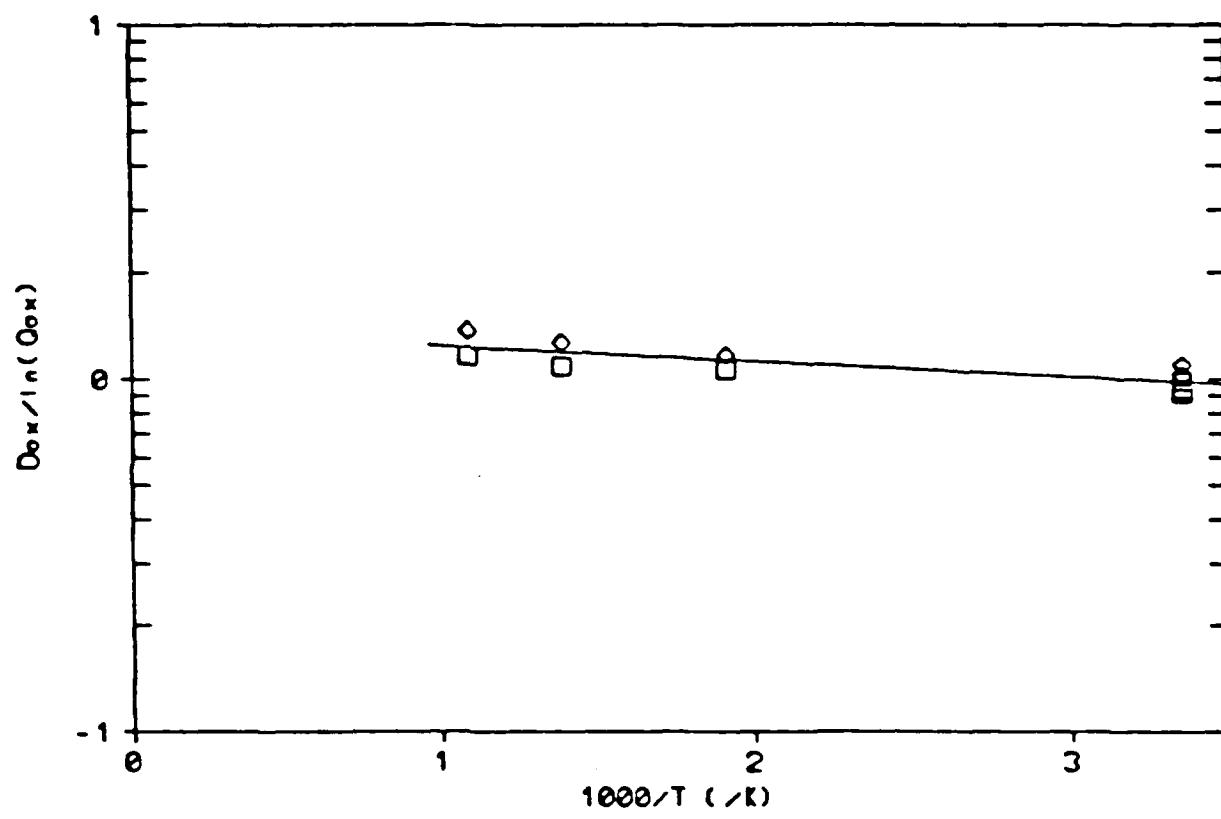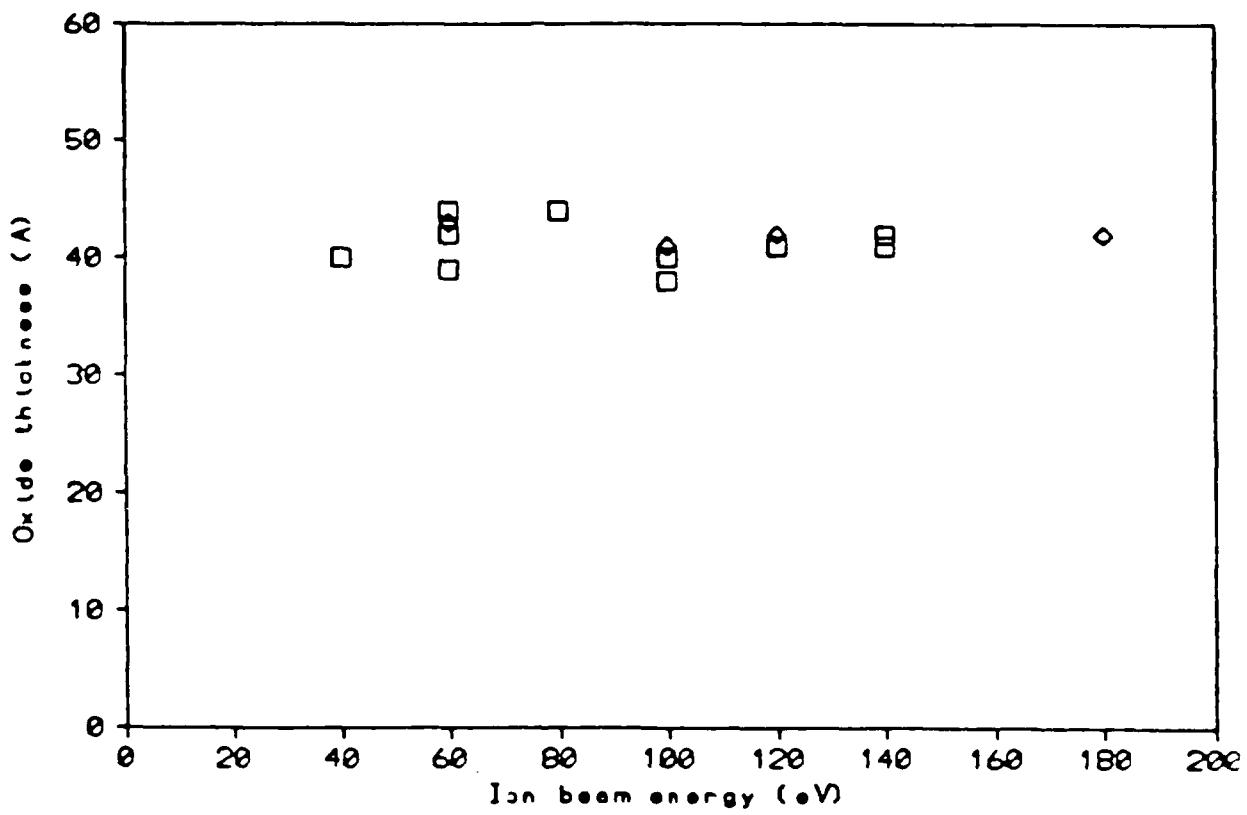
SPUTTERING OF SILICON DIOXIDE NEAR THRESHOLD

S.S. Todorov[a] and E.R. Fossum

Department of Electrical Engineering and

Columbia Radiation Laboratory

Columbia University

New York, New York 10027

Abstract

The sputtering rate for silicon dioxide by argon ion bombardment at energies appropriate for ion beam deposition (< 100 eV) has been measured. It has been found that the energy dependence of the oxide sputtering rate at these low energies is easily predicted by assuming the yield is limited by the metallic component of the binary target. This assumption is shown to also predict the sputtering rate of other metallic oxides.

[a]Permanent Address: Institute of Electronics, Sofia, 1184 Bulgaria

Bombardment of growing thin films by low energy ion beams has evolved into an important technique for the modification and deposition of thin films. Ion beam etching and reactive ion etching have achieved widespread application. More recently, the use of lower ion energies has led to the development of direct ion beam deposition applied to low-temperature epitaxial growth of Si and Ge[1-3]. Oxygen and nitrogen ion beams have been used for the controlled growth of ultra-thin oxide and nitride films on metal and semiconductor substrates[4-7].

These techniques allow good control over the composition, stoichiometry, orientation and crystallinity of the obtained films. The basic factor which limits the film growth rate in ion beam deposition is the sputtering due to the ion bombardment. The growth rate can be approximated by $dx/dt = (1/N)(\eta\bar{\Phi} - Y\bar{\Phi})$, where N is the density of the material, η is the sticking coefficient or incorporation probability, Y is the sputtering yield and $\bar{\Phi}$ is the ion flux. For deposition to take place it is necessary that $Y < \eta$. In the case of ion beam oxidation, it has been observed that a limiting oxide film thickness is reached which is related to the sputtering rate of the material[4,7]. It is evident that sputtering plays an important role in film growth by direct ion bombardment.

In this letter we report experimental results of sputtering of SiO_2 by Ar ions at energies below 100 eV. Accurate knowledge of the sputtering rate and its energy dependence at these low energies is becoming increasingly important in microelectronics device fabrication processes. The use of inert-ion bombardment precludes film growth and allows the separate investigation of the sputtering component of the deposition process. It is shown that a simple

- 2 -

assumption leads to good agreement between existing theory for non-binary materials and the reported measurements, not only for the case of silicon dioxide but also for other oxide targets.

The experiments are performed on samples prepared by oxidizing chemically cleaned p-type and n-type Si (100) wafers in dry oxygen at 900°C for times ranging between 10 min and 3 hrs and resulting oxide thicknesses of 45 Å to 520 Å. The thickness of the oxide films is measured with a He-Ne laser ellipsometer immediately before loading them into the ion beam processing chamber. The samples are mounted 15 cm from the ion source in a vacuum chamber with a base pressure of 2×10^{-7} torr. Broad Ar^+ ion beams with energies in the range 40-100 eV are produced by a single grid Kaufman ion source. The SiO_2 samples are bombarded with argon doses ranging from 1×10^{17} to 1.6×10^{18} cm^{-2}. The thickness of the remaining oxide is remeasured by ellipsometry after the samples are unloaded from the vacuum system. The accuracy of the ellipsometric data is approximately 1 Å, and although it may be distorted by damage to the substrate[8], this is not expected to be significant in this case. For example, measurements on substrates in which the oxide has been completely sputtered away indicate a "residual" film thickness of 3-5 Å, attributable to a monolayer of native oxide grown after the substrate is removed from the vacuum system.

The measured change in oxide thickness is plotted as a function of the ion dose for each value of ion energy. Six to ten different data points are plotted for each energy and the value of the sputtering yield is determined from the slope of the least-squares fit through the data points. The data for the case of 60 ev Ar^+ bombardment is shown in Fig. 1. The thickness decrease with

increasing ion dose is linear and independent of the initial oxide thickness.

The measurement of the sputtering rates at energies close to the sputtering threshold is difficult to measure due to the relatively low yields involved. For example, Oostra et al.[9] did not observe any sputtering of SiO_2 at argon energies less than 50 eV when examining scanning electron micrographs. It is believed that the method used here allows significantly more precise measurement of the sputtering rate, and material removal has been observed at energies as low as 40 eV.

The experimental results on sputtering SiO_2 by Ar ions are summarized by Fig. 2., where they are compared to the experimental data of Jorgenson and Wehner[10] and the higher energy data of Oostra et al.[9] Note that although widely cited, the data of Jorgenson and Wehner is obtained under quite different experimental conditions with the oxide film covering a thin tungsten wire immersed in a DC discharge.

Theoretically, the sputtering of monatomic targets is well understood[11] and good descriptions of the sputtering yield near threshold can be obtained by semi-empirical formulas[12-14] which introduce modifications of Sigmund's theory to account for the existence of a sputtering threshold. Very good agreement with experimental data is generally observed.

The sputtering process for multicomponent targets is more difficult to describe. The collision cascade process will affect each target component differently. Apart from causing the ejection of atoms from the target surface, the ion bombardment may also lead to target mixing, segregation, enhanced diffusion, and other

- 4 -

processes.[15] Thus, the correct description must take into account the changes in composition of the target brought about by the ion bombardment and how these changes are reflected in the sputtered particle flux. This discussion will concern only the sputtering of oxides; the behavior of compounds or alloys consisting only of low-vapor pressure materials may be quite different.[16]

In this letter, the simple assumption is made that oxygen is initially preferentially ejected and that the non-volatile component (silicon) limits the sputtering process. This assumption is supported by prior experimental evidence that the sputtering of metallic oxides by low energy (<1 keV) ions leads to the enrichment of the metal component within several monolayers of the surface[16]. This effect is expected to become especially pronounced at energies near the metal sputtering threshold, in agreement with the general observations on sputtering of binary targets whose components have different sputtering threshold energies.[15]

Further support for this assumption comes from previously reported experimental data concerning the sputtering yield of Ta and Ta_2O_5.[17-20] Our assumption correctly predicts that the removal of equal numbers of tantalum atoms leads to a 2.48-fold decrease in the volume of Ta_2O_5 as compared to Ta, for 100 eV Ar^+ and 500 eV He^+ bombardment. This ratio degrades with higher bombardment energies since the ejection of the material takes place from a layer whose composition can no longer be approximated as pure tantalum.

Applying this assumption to the case of low-energy argon sputtering of silicon dioxide leads to similarly correct predictions. The removal of the same number of silicon atoms will decrease the volume of a SiO_2 sample 2.21 times the change in the

- 5 -

volume of a silicon sample. Thus, using the sputtering yield of silicon calculated according to Matsunami et al.[13] with the threshold energy E_{th} = 27 eV as suggested by Yamamura and Bohdansky[14] leads to excellent agreement with the experimentally obtained data, as shown in Fig. 3. The inadequacy of the assumption at higher energies is pointed to by the slight deviation of the data from the predicted curve.

In summary, the sputtering yield of SiO_2 under low energy argon ion bombardment has been measured in the range 40-100 eV. An assumption for understanding the energy dependence in terms of low energy silicon sputtering with a simple volumetric rate adjustment leads to excellent prediction of the experimental data. The assumption has been extended to predicting other oxide sputtering rates with equal success.

1. J. Amano, P. Bryce, and R.P.W. Lawson, J. Vac. Sci. Technol., 13, 591 (1976).

2. G.E. Thomas, L.J. Beckers, J.J. Vrakking, and B.R. de Koning, J. Crystal Growth, 56, 557 (1982).

3. N. Herbots, B.R. Appleton, T.S. Noggle, R.A. Zuhr, and S.J. Pennycook, Nucl. Instr. Meth., B 13, 250 (1986).

4. J.M.E. Harper, M. Heiblum, J.L. Speidell, and J.J. Cuomo, J. Appl. Phys., 52, 4118 (1981).

5. S.S. Todorov, S.L. Shillinger, and E.R. Fossum, IEEE Electron Device Lett., EDL-7, 468 (1986).

6. C.F. Yu, S.S. Todorov, and E.R. Fossum, J. Vac. Sci. Technol., A-5, 1569 (1987).

7. S.S. Todorov and E.R. Fossum, to be published in J. Vac. Sci. Technol.

8. D.E. Aspnes and A.A. Studna, Surf. Sci., 96, 294 (1980).

9. D.J. Oostra, R.P. van Ingen, A. Haring, A.E. de Vries, and G.N.A. van Veen, Appl. Phys. Lett., 50, 1506 (1987).

10. G.V. Jorgenson and G.K. Wehner, J. Appl. Phys., 36, 2672 (1965).

11. P. Sigmund, Phys. Rev., 184, 184 (1969).

12. J. Bohdansky, Nucl. Instr. Meth, B-2, 587 (1984).

13. N. Matsunami, Y. Yamamura, Y. Itikawa, N. Itoh, Y. Kazumata, S. Miyagawa, K. Morita, R. Shimizu, and H. Tawara, Atomic Data and Nuclear Data Tables, 31, 1 (1984).

14. Y. Yamamura and J. Bohdansky, Vacuum, 3, 561 (1985).

15. H.H. Andersen, In: Ion Implantation and Beam Processing, ed. J.S. Williams and J.M. Poate, Acad. Press, New York, 1984.

16. G. Betz and G.K. Wehner, In: Sputtering by Particle Bombardment II, ed. R. Behrisch, Springer-Verlag, Berlin, 1983.

17. H. v. Seefeld, R. Behrisch, B.M.U. Scherzer, P. Staib, and H. Schmidl, In: Proc. VII Int. Conf. Atomic Collisions in Solids, Moscow, 1977, Vol. 2, 327. Moscow State Univ. Publ. House, 1980.

18. N. Laegried and G.K. Wehner, J. Appl. Phys., $\underline{32}$, 365 (1961).

19. H. Oechsner, H. Schoof, and E. Stumpe, In: Proc. 7th Intern. Vacuum Congress and 3rd Intern. Conf. Solid Surfaces, ed. R. Dobrozemsky, F. Rudenauer, F.P. Viehbock, and A. Breth, Wien, 1977.

20. J. Roth, J. Bohdansky, and W. Ottenberger, Report IPP 9/26, Max-Planck-Institut fur Plasmaphysik, Garching, FRG, 1979.

Figure Captions

Fig.1.   Measured decrease in the thickness of the SiO_2 film as a
          function of Ar dose.   E = 60 eV, Ar^+ flux = 2.2×10^{15}
          $cm^{-2}sec^{-1}$.

Fig.2.   Experimental results on sputtering of silicon dioxide by low
          energy argon ions.

Fig.3.   Comparison of calculated and measured sputtering rates for
          Ar^+-bombarded SiO_2.
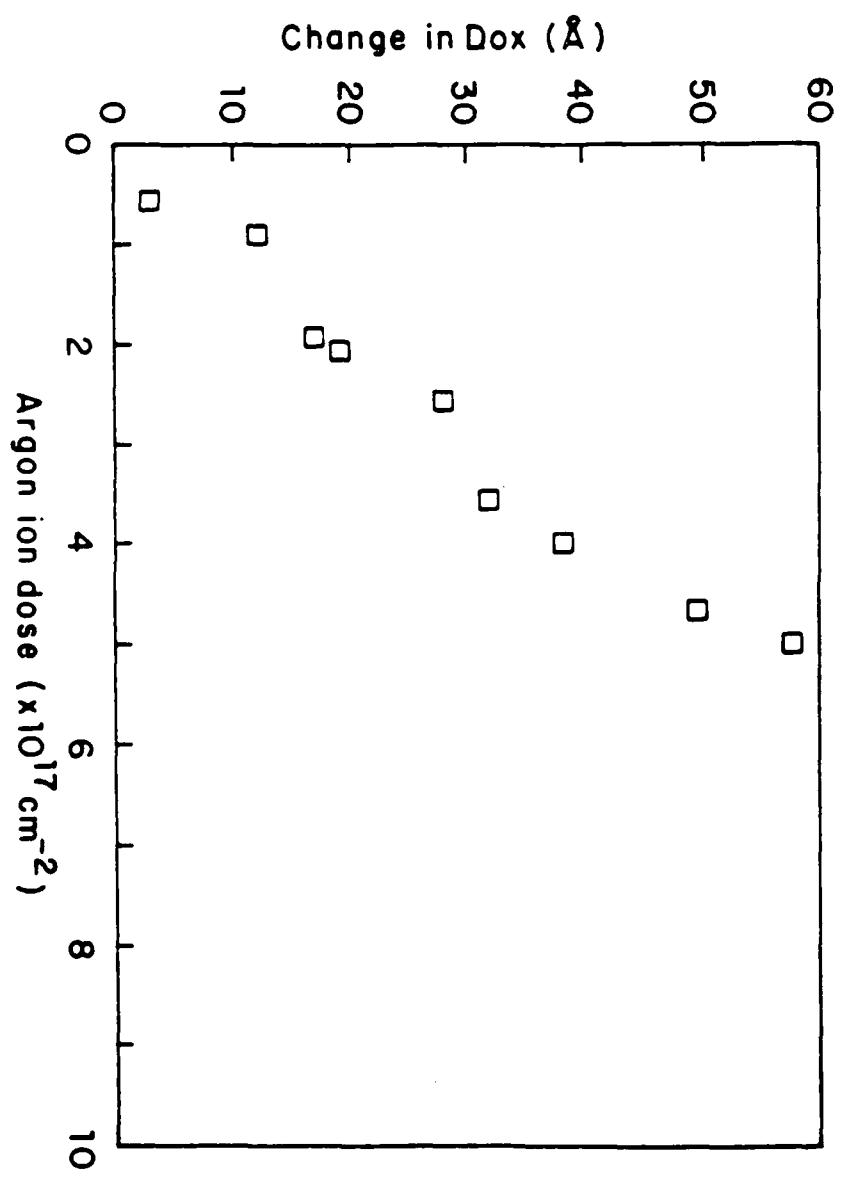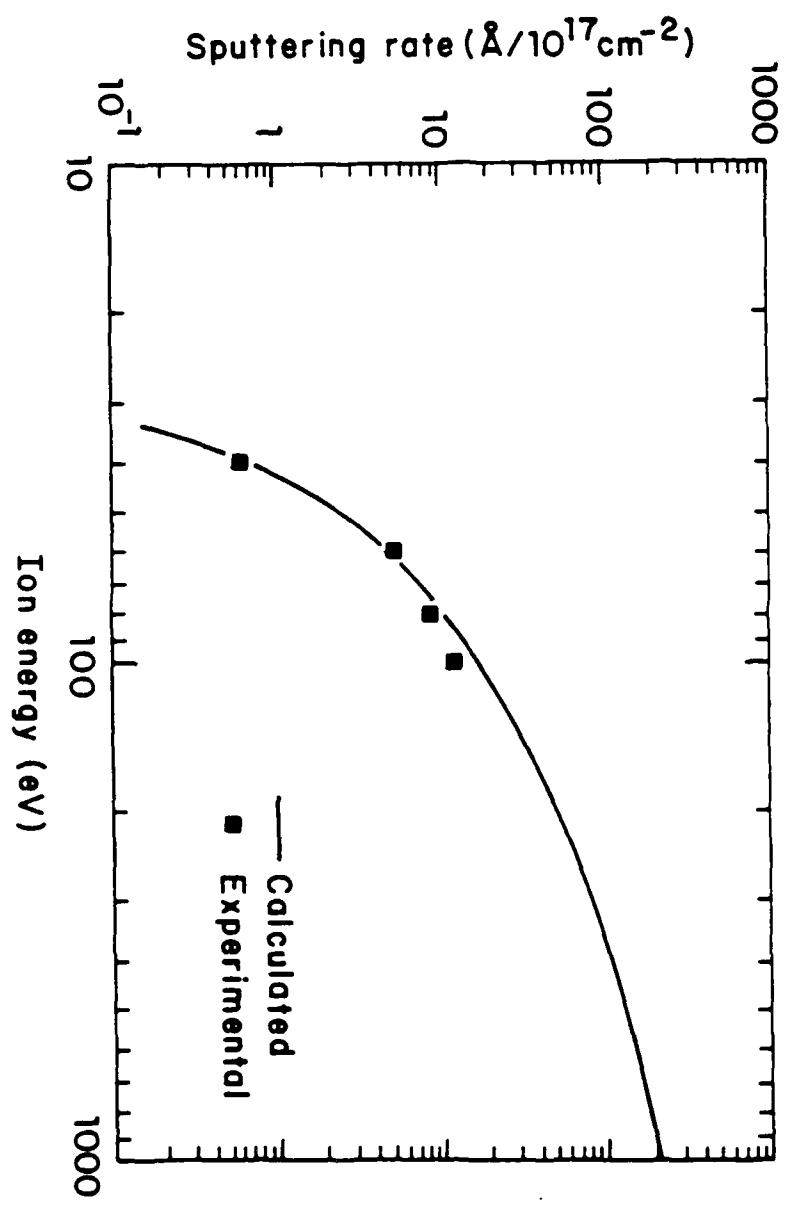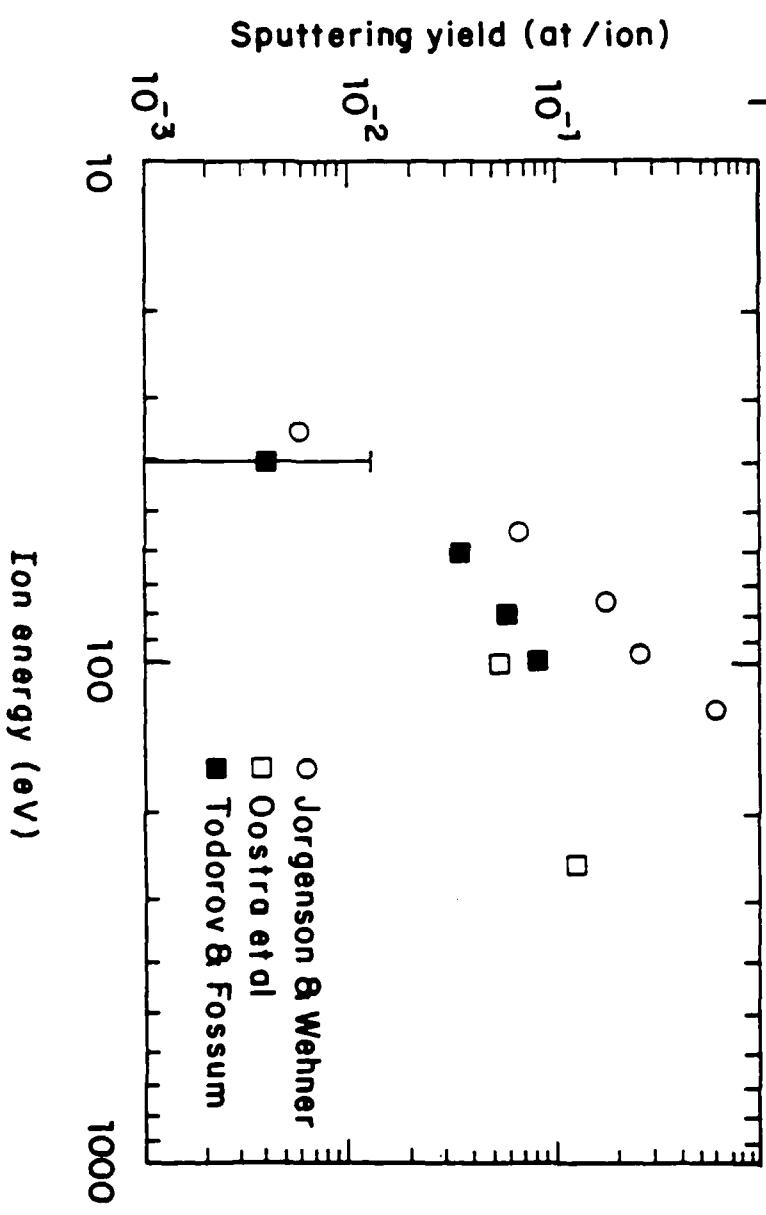
Figure: Sputtering yield (at/ion) versus Ion energy (eV).

Legend:
○ Jorgenson & Wehner
□ Oostra et al
■ Todorov & Fossum

OXIDATION OF SILICON BY A LOW ENERGY ION BEAM:

EXPERIMENT AND MODEL

Eric R. Fossum

S.S. Todorov and E.R. Fossum

Department of Electrical Engineering

and MIcroelectronics Sciences Laboratories

Columbia University

New York, New York   10027

Abstract

The self-limiting oxidation of silicon by a low energy ion beam
(40-120 eV) is described by an implantation-sputtering model. The
thin oxide (40-50 Å) is grown primarily by a surface implantation
process which leads to a logarithmic increase of oxide thickness with
dose in the absence of sputtering. At higher energies (100 eV), the
sputtering of the growing film leads to net self-limiting growth.
The model, which does not include adjustable parameters, is used to
describe the dose evolution of the oxide growth as a function of beam
energy. The implantation-sputtering model is found to be in
excellent agreement with experimental observations.

Eric R. Fossum

The room temperature oxidation of silicon by the use of a broad, low-energy ion beam containing a mixture of argon and oxygen has been shown to produce an oxide suitable for use as a gate dielectric in metal-oxide-semiconductor (MOS) transistors.[1,2] The ion beam energy is typically in the range of 40-120 eV with doses of the order $10^{17}/cm^2$ delivered in 3-5 minutes, leading to oxides of thickness 40-50 $\mathring{A}$. The oxides are found to be stoichiometric SiO_2 as measured by Auger electron spectroscopy and x-ray photoelectron spectroscopy. The thickness of the films is measured after growth using several techniques - the primary one being ellipsometry[3]. The film thickness has a weak dependence on substrate temperature in the range $25^\circ C$-$650^\circ C$, on dose, and beam energy.[4] The oxide thickness uniformity as measured by test sites across a wafer is excellent and far exceeds the dose uniformity. This letter presents an implantation-sputtering model to describe the oxidation process. The model predictions are in excellent agreement with the above experimental observations.

In the implantation-sputtering (I-S) model, the oxide growth is governed by a surface implantation process, rather than by thermal diffusion. This is in contrast to a previous model suggested for the low-energy ion beam oxidation of metals.[5] The range distribution of the oxygen ions in the target is calculated using linear cascade theory.[6,7] Energy loss occurs entirely by elastic binary collisions with the target nuclei. The target is treated as homogeneous, isotropic and monatomic with target-atom mass and atomic number determined according to Bragg's rule and taking into account the gradual transformation of the target from silicon to silicon

dioxide. The interatomic potential used is the universal potential.[8] Power-potential approximations[7] obtained by a least-squares fit to this potential are used to allow analytical solution. Thus, m=1/5 is used to fit the potential at reduced energy in the range $10^{-5}-2\times10^{-3}$. The mean and the straggling of the distribution are calculated and it is fitted by a simple Gaussian. The results of this calculation in the case of 50 eV O bombardment of SiO_2 are presented in Fig. 1. This corresponds to the case of 100 eV O_2 ion beam bombardment. It should be noted that while the mean range is predictably small (approximately 5 Å) the straggling is significant. It is this straggling tail which governs the oxide growth evolution.

In the I-S model, the oxide growth competes with the simultaneous sputtering of the oxide film by the same beam. Since the experiments are primarily performed with 50%-50% argon-oxygen mixtures, the sputtering of the target is due mainly to the argon bombardment. The sputtering rate of SiO_2 under low energy argon ion bombardment has been measured in a separate set of experiments[9] and the results are shown in Fig. 2. Sputtering at 60 eV can be considered negligible for the oxide thicknesses considered, but becomes appreciable by 100 eV.

A computer simulation of the oxide growth evolution has been performed using the I-S model. The incorporation of oxygen into the target and the conversion of Si to SiO_2 is accompanied by volume expansion. To account for this, the evolution of the oxide layer is calculated in small time (oxygen dose) increments. The range distribution is recalculated for each time interval. It is assumed

that the implanted oxygen reacts immediately with unreacted silicon, since the former may be considered "hot" in the last stages of stopping. However, oxygen coming to rest within stoichiometric SiO_2 is assumed to outdiffuse to the surface. In thicker oxides, consideration of a spatially dependent diffusion coefficient may become necessary in view of the depth dependence of the beam energy deposition and consequent defect production.

The stoichiometry of the film as a function of depth for increasing dose is shown in Fig. 3 for the case of 100 eV O_2 bombardment. An interesting feature of the evolution is apparent. It is seen that the interface between oxide and silicon is initially broad, but as the surface erodes due to sputtering, the interface becomes more sharp. The width of the interface explains electrical characterization results, such as lower channel mobility in ion beam oxide MOS transistors, and anomalously larger capacitance-voltage measurements of oxide thickness.[10]

The simulated dependence of the oxide film thickness on ion dose for 100 eV bombardment is presented in Fig. 4 where it is compared to experimental results. Since the experimental data is obtained by ellipsometry, the simulated thickness is determined in two ways. In the low dose case where no stoichiometric oxide has been formed, the thickness is plotted according the position of the steepest gradient in oxygen concentration, which is approximately halfway through the transition region. For higher doses, in which stoichiometric SiO_2 has been formed, the edge of the stoichiometric region is used to assess the thickness. For 60 eV bombardment simulations in which sputtering is negligible, the film thickness grows logarithmically

with dose, and is also in excellent agreement with experimental observations.

The simulated final oxide thickness as a function of ion beam energy is shown in Fig. 5. Also shown is the experimental data, which shows good agreement. Note that the experimental and simulated data show a maximum in thickness for the 50-60 eV energy range. The maximum is a result of the competing processes of implantation and sputtering. The calculated thickness for 140 eV bombardment is lower than the experimental data due to an overestimation of sputtering in the I-S model. The low energy ion beam sputtering model of oxides assumes that sputtering of oxygen is not rate limiting - an assumption which begins to fail at higher energies.[9]

In summary, an implantation-sputtering model without adjustable parameters has been presented for the low energy ion beam oxidation of silicon. The model explains the weak dependence of oxide thickness on temperature, energy, and dose. The excellent agreement between simulation and experiment allows the important conclusion that linear cascade theory and the binary collision approximation continue to be applicable at these low ion energies.

REFERENCES

1. S.S. Todorov, S.L. Shillinger, and E.R. Fossum, IEEE Electron
    Device Lett., EDL-7, 468 (1986).

2. S.S. Todorov and E.R. Fossum, Proc. MRS Symp., Boston, Dec. 1986.

3. C-F Yu, S.S. Todorov, and E.R. Fossum, J. Vac. Sci. Technol.,
    A-5, 1569 (1987).

4. S.S. Todorov and E.R. Fossum, presented at the 31st Electron,
    Ion, and Photon Beam Conference, Woodland Hills, CA, May 1987,
    and to be published in J. Vac. Sci. Technol.

5. J.M.E. Harper, M. Heiblum, J.L. Speidell, and J.J. Cuomo, J.
    Appl. Phys., 58, 4118 (1981).

6. J. Lindhard, M. Scharff, and H.E. Schiott, Kgl. Dan. Vidensk.
    Selsk. Mat. Fys. Medd., 33, No. 14 (1963).

7. K.B. Winterbon, P. Sigmund, and J. Sanders, Kgl. Dan. Vidensk.
    Selsk. Mat. Fys. Medd., 40, No. 14 (1970).

8. J.F. Ziegler, in Ion Implantation Science and Technology, ed.
    J.F. Ziegler, Academic Press, Orlando, 1984.

9. S.S. Todorov and E.R. Fossum, submitted to Appl. Phys. Lett.,
    July 1987

10. S.S. Todorov, Ph.D. Thesis, Columbia University, 1987.

FIGURE CAPTIONS

1. Normalized range and deposited energy distributions calculated
    for 50 eV O^+ on SiO_2.

2. Sputtering rate for low energy argon ion bombardment of silicon
    dioxide. Rate is measured in angstroms sputtered per dose of
    10^{17} incident ions per cm^2.

3. Simulation of oxide growth evolution. Plot shows bonded oxygen
    concentration as a function of depth for 100 eV O_2^+ ion
    bombardment. Numbers refer to increasing doses.

4. Oxide thickness as a function of dose measured experimentally and
    simulated using the implantation-sputtering model for 100 eV
    O_2^+ bombardment.

5. Oxide thickness as a function of ion energy measured
    experimentally and simulated using the implantation-simulation
    model. O_2 dose: 3×10^{17} cm^{-2}.

Range
Deposited energy

Depth in target (Å)

Fig. 5-1b

SiO$_2$

Depth into target (Å)
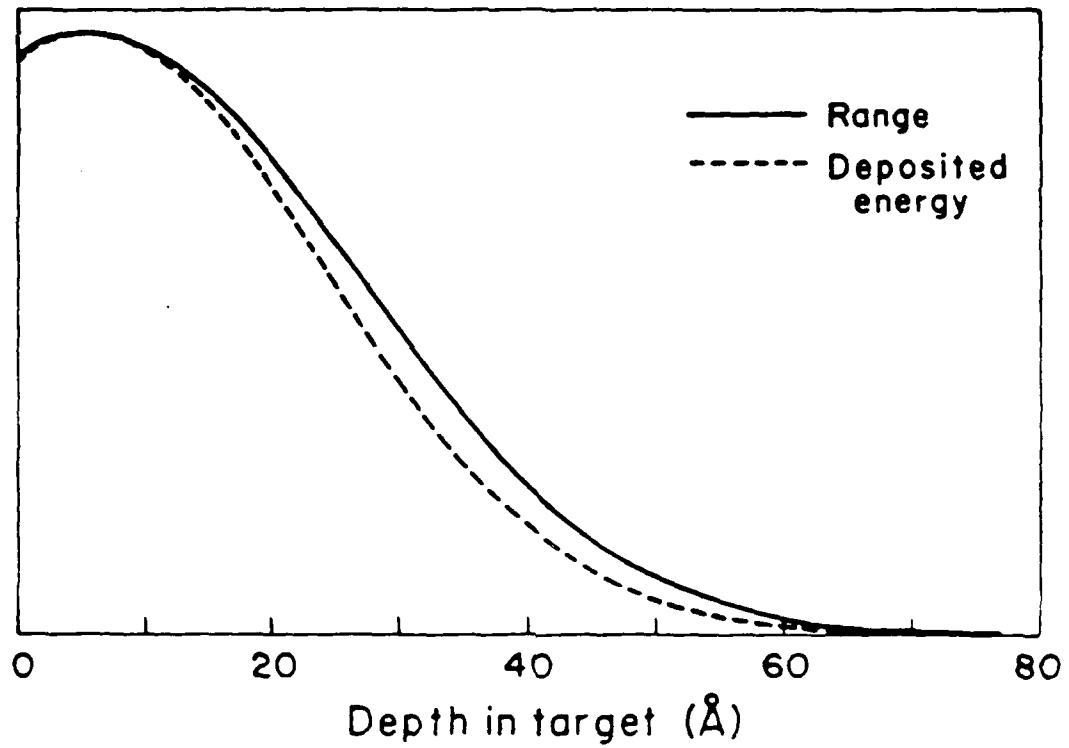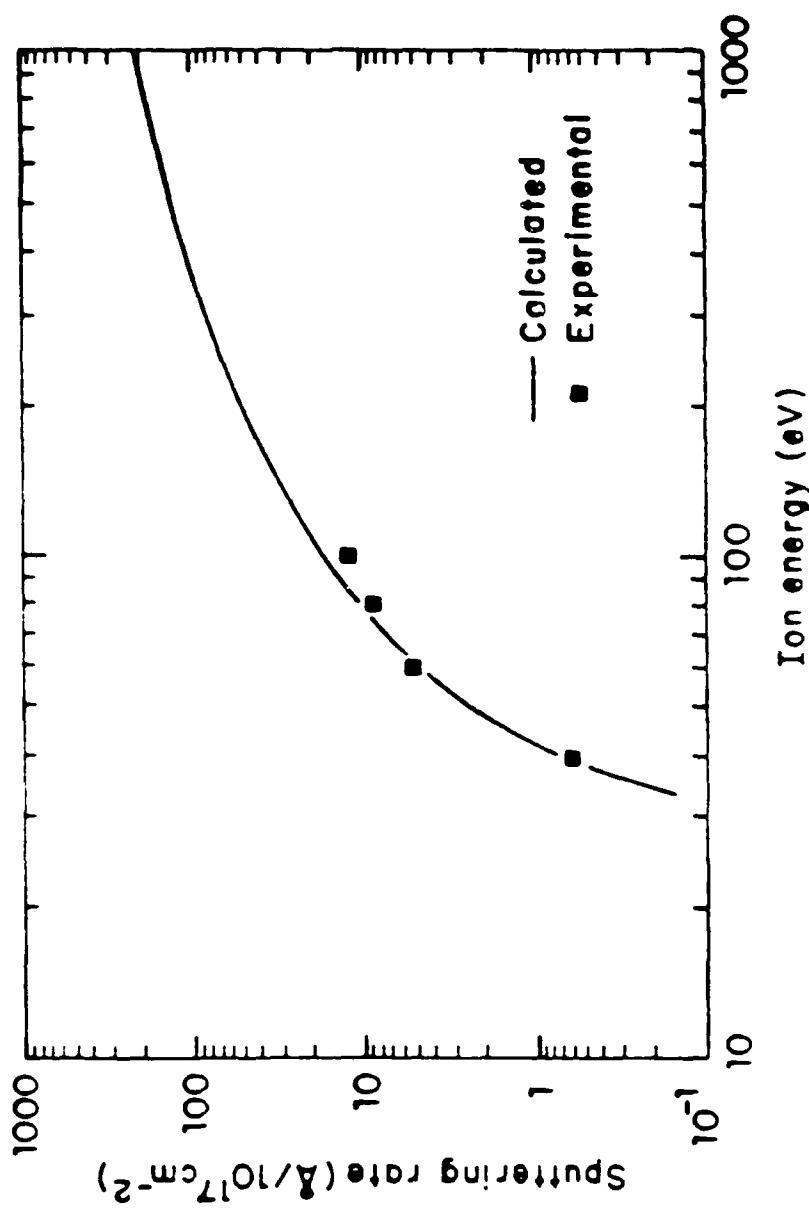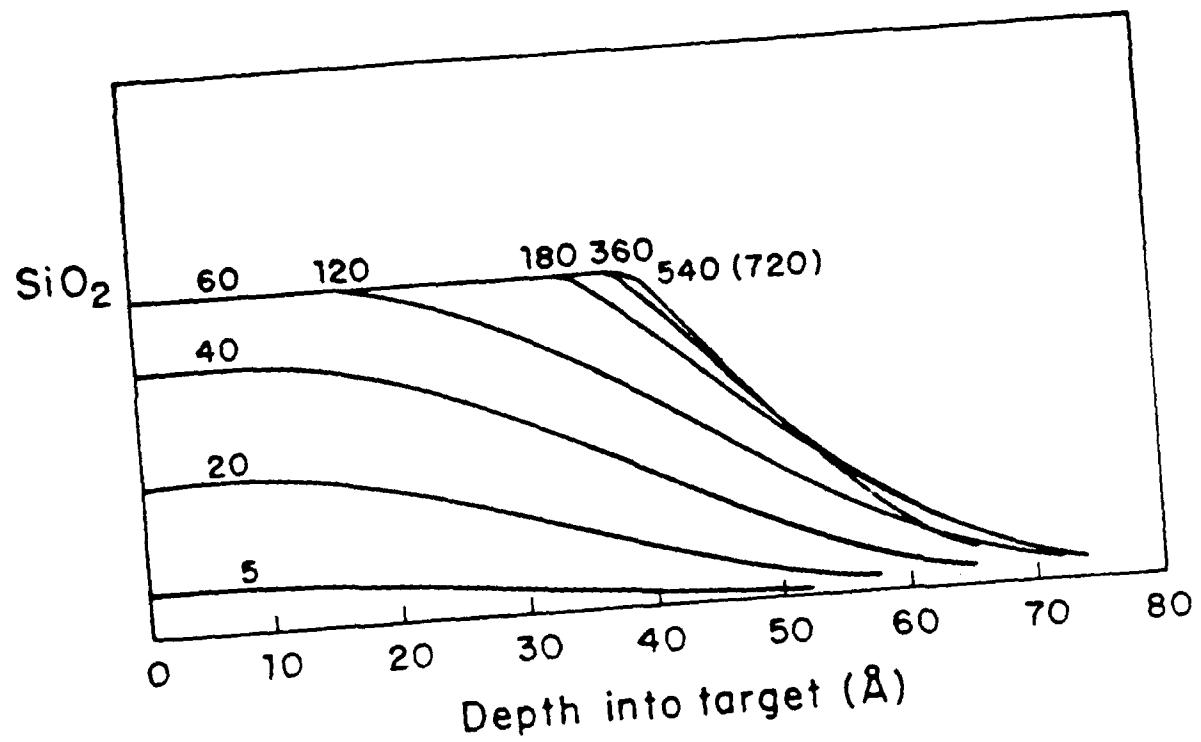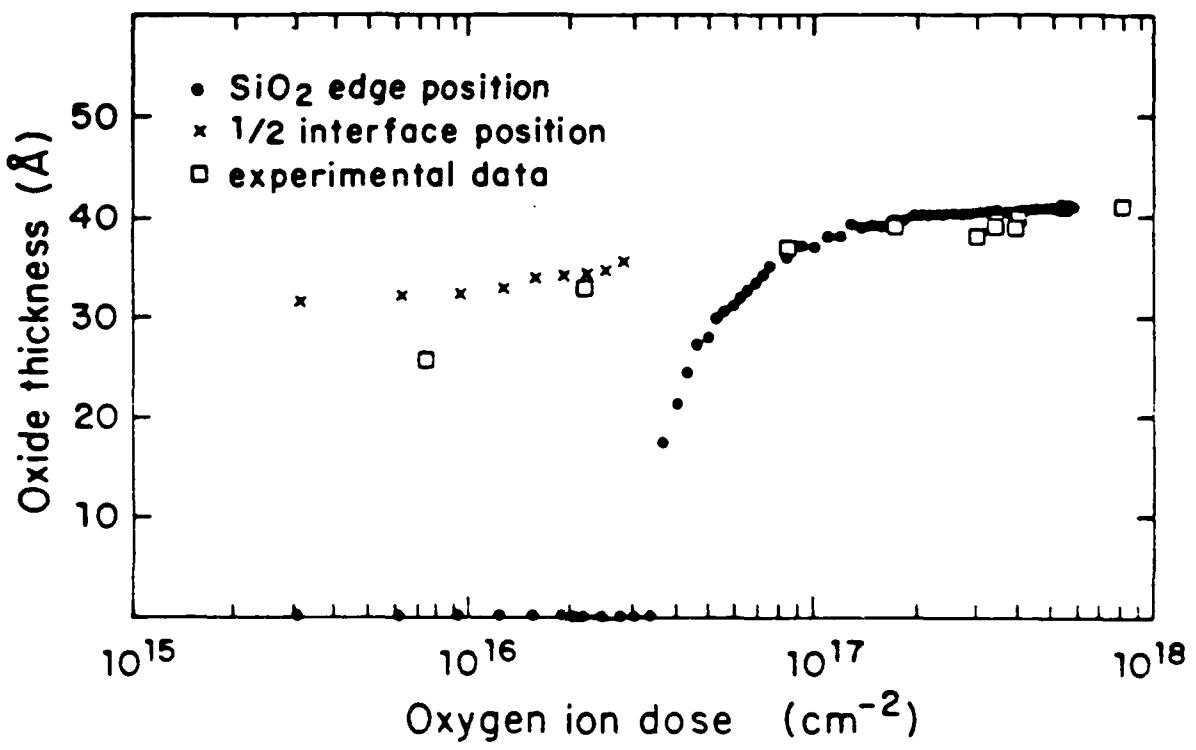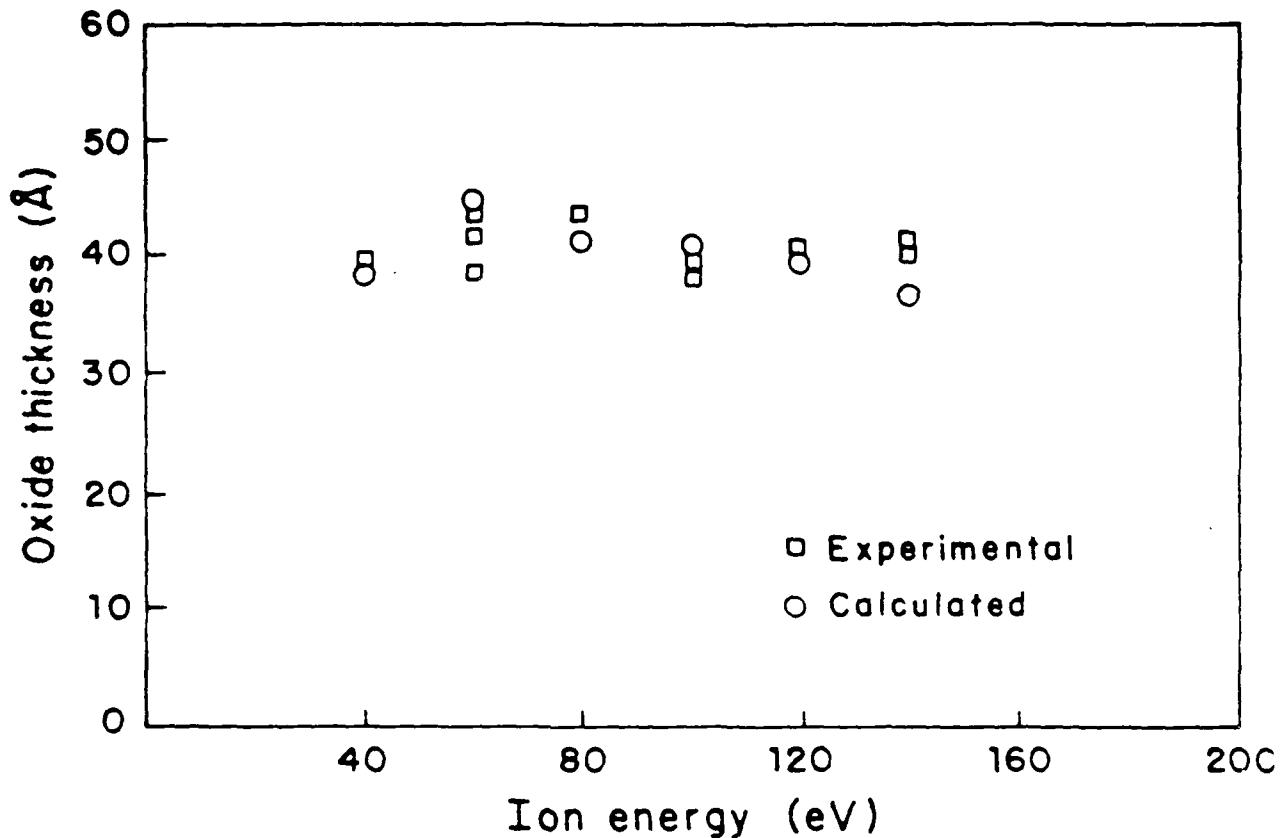
INTERACTIONS OF LOW ENERGY
OXYGEN IONS
WITH
SILICON SURFACES

Stanislav Stefanov Todorov

COLUMBIA UNIVERSITY

1987

USING INCOHERENT LIGHT TO GENERATE COHERENT EXCITATIONS

S. R. Hartmann
Physics Department, Columbia University, New York, N.Y.

ABSTRACT

Photon echo generation by incoherent light is both practical and efficient. In the regime where the noisy excitation pulses overlap fast transient response is obtained and in addition psec beating is observed in Na vapor, the latter arising from coherent superposition states associated with the D line transitions. Only the fast response requires broadband noisy light: ultrafast modulated echoes are produced with either broadband noisy light spanning the D line transitions or with a pair of narrow band lasers resonant with each of the D line transitions.

INTRODUCTION

The photon echo effect is normally thought of as a coherent effect brought about by the interaction of an ensemble of atoms or molecules with a sequence of laser pulses which must be coherent.[1] This notion is not altogether correct as was demonstrated in '84 when an echo experiment was performed with the usual coherent laser sources replaced by broadband amplified spontaneous emission sources.[2] Not only was it possible to generate echoes in this unorthodox manner but the echoes so produced were larger than those obtained using coherent laser pulses.

Although usually not explicitly stated it is understood that the laser excitation pulses must be resonant with the optical transition to be excited. A fuller appreciation of this obvious requirement leads to an understanding of the unexpected effectiveness of echo generation with noise. With this motivation we ask the question: What constitutes resonant radiation? For the case in which all homogeneous relaxation times are long compared to the laser pulse width and the lasers are weak so that power broadening is negligible, then all atoms within a bandwidth $\Delta\omega = 1/\tau_p$, where τ_p is the duration of a coherent laser pulse at frequency ω, interact strongly with the radiation field. Outside this regime the interaction falls quickly and can therefore be called nonresonant. A stronger statement than this can in fact be made which is that the interaction strength within this bandwidth is uniform. That is, an atom at exact resonance and one removed by $\Delta\omega$ from exact resonance experience a pulse with the same area θ. But our interest is with incoherent radiation pulses. We now note that if such pulses have a duration τ_p then any radiation sampled for that duration and lying within a bandwidth no greater than $\Delta\omega = 1/\tau_p$ from exact resonance must appear coherent. This is an uncertainty argument result. It follows that any particular atom sees the applied incoherent laser field as a coherent one. It also follows that all atoms within the common bandwidth $\Delta\omega$ see the same coherent field and behave as a coherent group. With these results it is clear that echo production with incoherent light is to be expected. Each group of atoms within a bandwidth $\Delta\omega$ acts as a unit. If P_{echo} is the coherent dipole moment density of one such unit the radiated intensity I_{echo} of the atom ensemble varies as $N P_{echo}^2$ where N is $\Delta\Omega/\Delta\omega$, the number of units within the linewidth $\Delta\Omega$.

Echo experiments on the Na D lines in Na vapor using single mode coherent 7 nanosecond laser pulses can only excite the fraction $\Delta\omega/\Delta\Omega \approx 1/30$ of the resonance line. Incoherent echo generation gains us a factor of 30 by utilizing all the atoms in the resonance line. The statistical nature of the incoherent light leads to a reduction to 33% in the average value of P_{echo}^2 but the overall advantage of using incoherent light in this situation still corresponds to a factor of $(1/3)(30)$ or 10. In the echo experiments we had performed using

coherent light the lasers ran in several modes and so we only achieved an enhancement factor of 3 when incoherent sources were used. We constructed the incoherent sources by taking ordinary dye lasers and removing all elements which might define a cavity. Over the region where these incoherent sources lased, their output was effectively thermal.

FOUR WAVE MIXING

The object of these experiments was to establish that echo generation by incoherent light was both possible and practical. As an adjunct to this work a four wave mixing signal was developed in Na vapor at a temperature well above 445 K with temporally overlapping pulses from a single intense incoherent laser (thermal) source. The source output was split to produce two separate pulses we label 1 and 2 and then recombined after introducing an optical delay. The result is shown in figure 1.[1] When the delay between pulse 1 along k and 2 along k + K exceeds the 7nsec duration of these pulses the resulting four wave mixing signal along k + 2K is just the ordinary photon echo. The data we show here goes only to the 7nsec limit. The purpose of this experiment was to show that no well defined laser modes existed in our amplified spontaneous emission laser source. Had they been present there would have been a beat structure on the four wave mixing signal. The rapid fall in intensity for negative τ occurs in the order of a 100 psec. For positive τ there is a signal fall off with a time constant of the order of a few hundred psec followed by a slower fall off due to the diminishing of the temporal overlap of the excitation pulses.


FIG. 1 Four-wave mixed signal, generated by use of one thermal source, as a function of pulse separation. The absence of any large-scale beating demonstrates that there is no stable mode structure in the thermal excita-tion pulses.

At the time of this work parallel experiments were being carried out by Yajima et al[3] and Asaka et al[4]. The time delayed four wave mixing experiments of Yajima et al. used picosecond noise pulses and the experiments of Asaka et al. used a noisy cw laser. In both these cases the laser excitations were weak and a perturbation theory was suitable.

In an effort to understand better the interaction of noisy light with matter we proceeded to study four wave mixing with intense radiation fields.[5] We further broadened the spectral width of our incoherent source, and reduced the temperature of the Na vapor to 445 K so that collisional broadening effects were negligible. We found that as had been reported by Yajima we obtained beats in the four wave mixing signal when we varied the time delay τ. These beats correspond to the 6 Å splitting of the Na D lines.[6] Whereas our incoherent echo experiments were performed using light with a bandwidth of only a few GHz our four wave mixing experiments with intense fields had a bandwidth of 12 Å, sufficient to excite both D lines. Our results are shown in figure 2.[5] These results are similar to those of figure 1 except for the dramatic 1.9 psec beating. The fast fall off in intensity S for positive τ wasn't expected as we now worked in a collision free regime where the result of Morita and Yajima[7], valid when the longitudinal and transverse relaxation times, T_1 and T_2, are large compared to the inverse 1/e half width of the inhomogeneous line, $\delta\omega^{-1}$, applies, as was the case in our
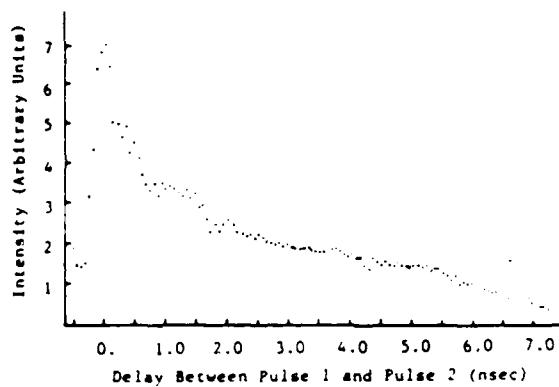
$$S = \int_{-\delta\omega\tau/\sqrt{2}}^{\infty} dy \, exp(-y^2)$$

experiment. Our observed degradation in S for large τ is due to laser noise induced relaxation effects and is a subject of continuing interest.[8] Our discussion here will be limited to explaining why coherent signals should be produced by broadband noise and to what extent the broadband noise is essential in effecting a particular result. We limit ourselves to the weak excitation limit and we refer the reader to the work of Morita and Yajima for an analytic analysis of many aspects of this problem.

SIMPLE ANALYSIS

First we present a simple picture of the time delayed four wave mixing process. We begin by modeling the incoherent laser excitation as a series of delta function spikes randomly phased. The laser output is divided and recombined so that the sample is irradiated with two identical laser pulses directed along k and $k + K$ with relative delay τ. These two spike trains are pictured in the upper part of figure 3. The lower part of this figure shows two recoil diagrams[9]. In the upper recoil diagram the spike along k preceeds the spike along $k + K$ while in the lower recoil diagram it follows it. In all parts of the figure the time variable is synchronized and increases to the

FIG. 2. (a) Experimentally measured signal intensity vs excitation pulse separation in a 10-cm-long Na cell at 445 K. (b) The same data for pulse separations of 50—100 psec with error bars and lines connecting the data. The 1.9-psec modulation is due to the Na $3P$ fine-structure beating.

right. Let's consider the upper recoil diagram. It shows the effect of the first four noise spikes along k separately generating excited state amplitudes which recoil until deexcited by the corresponding noise spikes along $k + K$. In this manner a macroscopic ground state amplitude with momentum corresponding to $-K$ is built up. The essential thing here is that although the noise spikes in each directed pulse train are randomly phased there is a fixed phase difference between noise spikes separated by τ in the two pulse trains. When we consider the effect of a noise spike along $k + K$ acting on the ground state amplitude the effect is as shown in the central section of the upper recoil diagram. An excited state amplitude is generated which recoils along $k + K$ and collides τ later with the ground state amplitude recoiling along $-K$. The result is the formation of a macroscopic dipole moment with wave vector $k + K - (-K) =$

FIG. 3. Upper section: Two identical laser spike trains with relative delay τ along $k+K$ and k. Lower section: Recoil diagrams showing signal generation along $k+2K$ for positive τ (above) and for negative τ (below). Time increases to the right in all cases.
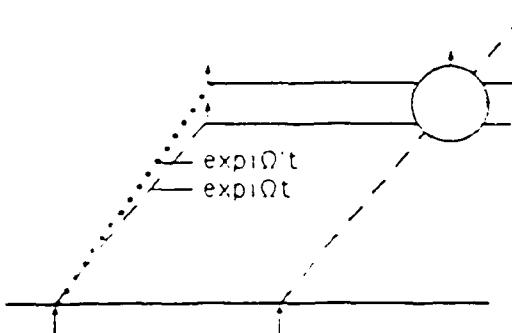

FIG. 4. Recoil diagram showing origin of ground superposition state modulation.

$k + 2K$ in the phase matched direction in which we look. Each noise spike along $k + K$ therefore generates a macroscopic dipole moment which radiates. Since these noise spikes are randomly phased the signal we see is the sum of the intensities which would be radiated separately by each of the induced dipole moment densities. The picture we have presented corresponds to a series of accumulated photon echoes which combine incoherently to yield the result expressed above in Eq(1). The signal falls to half its assymptotic limit at $\tau = 0$ since the echo amplitude is fully formed on application of the noise spike along $k + K$ and the contribution to the time integral of the echo signal during its formation is not obtained. For negative τ a similar analysis using the lower recoil diagram obtains and the resulting signal is seen to be best interpreted as a series of incompleted free decays.

The 1.9 psec beats we observed can be explained in terms of the macroscopic ground state population recoiling along $-K$ being modulated at the difference frequency of the sodium D lines. Consider figure 4 which shows the recoil diagram appropriate when two separate transitions are excited by the same delta function spike. Each excited state amplitude evolves according to its energy eigenvalue with the result that the ground state formed τ later has an amplitude proportional to $[\exp(-i\Omega\tau) + \exp(-i\Omega'\tau)] \sim \cos[(\Omega-\Omega')\tau/2]$ where Ω and Ω' are the eigenfrequencies associated with the two excited states. As before, the effect of a later spike along $k + K$ is to generate a recoiling state which collides with the reformed ground state to produce an echo. This ground state

having an amplitude which is modulated at half the difference frequency of the two excited states gives rise to an echo intensity modulated by $\cos[(\Omega-\Omega')\tau]$. As can be seen from the figure this is true whether the interrogation spike which produces the recoiling excited state (states) is resonant with either of the two transitions or both.

The virtue of the recoil diagram approach is that it provides a graphic picture of the echo formation or time delayed four wave mixing process. If relaxation processes are introduced then the radiated signal will diminish with increasing τ because the superposition states will dephase reducing the echo formation directly as well as indirectly through an incomplete formation of the macroscopic ground state. It appears that this technique may be well suited for the measurement of fast relaxation phenomena.[10]

ULTRAFAST SPECTROSCOPY WITH NARROW BAND LASERS

The incoherent echo analysis was motivated by the realization that only radiation resonant with a transition will interact with it. This observation suggests an alternative method for generating the psec beats observed above. The method is to replace the broadband incoherent laser whose spectral output covers both lines with two narrowband coherent lasers each resonant with one of the transitions.[11] The output of these lasers is combined and treated as a single pulse. We performed a regular two pulse photon experiment using regular dye lasers with 7 nsec duration pulses. As expected we found that echoes are produced which are in all respects similar to echoes previously produced except that when we vary the pulse separation τ the overall echo intensity varies according to $\cos[(\Omega-\Omega')\tau]$. Thus the psec beating is obtained with narrow band excitation. This beating is observed even when a spectrometer is introduced in front of the detector so that radiation is detected from only one of the two transitions excited. This demonstrates that the beating arises from an interference in the common ground state population.
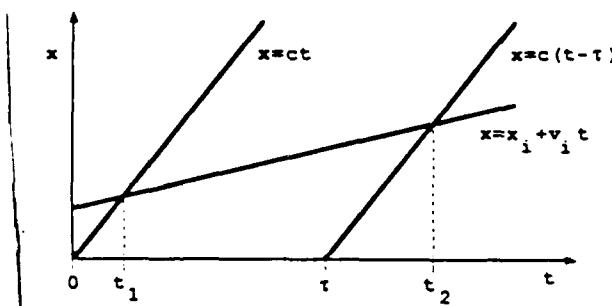

FIG 5 Space-time diagram showing excitation pulses and an atom traveling at v_i.

This result holds independent of the frequency splitting of the two optical transitions excited. By proper selection of materials narrow band excitation of beats in the femtosecond and attosecond regimes is possible. These beats can be observed with slow detectors as the beat is present in the integrated echo intensity signal not in the individual echo signal envelope.

The echo intensity is calculated by following a moving atom which is irradiated at t_1 and t_2. The lasers are not at exact resonance but must both excite the same velocity subgroup as we are interested only in the situation where beats occur. We use a + and – subscript to indicate time just before and after the pulse arrival and we list the wavefunction development as follows:

$t_{1-}> = |g>$, $t_{1+}> = |g> + |e> + |e'>$,

$t > t_{1+}> = |g> + \exp[-i\Omega(t-t_1)]|e> + \exp[-i\Omega'(t-t_1)]|e'>$,

$t_{2-}> = |g> + \exp[-i\Omega(t_2-t_1)]|e> + \exp[-i\Omega'(t_2-t_1)]|e'>$,

$t_2+\rangle = |e\rangle + |e'\rangle + \{\exp[-i\Omega(t_2-t_1)]|g\rangle + \exp[-i\Omega'(t_2-t_1)]|g'\rangle\}$,

$t\rangle t_2+\rangle = \exp[-i\Omega(t-t_2)]|e\rangle + \exp[-i\Omega'(t-t_2)]|e'\rangle$
$\quad + \{\exp[-i\Omega(t_2-t_1)] + \exp[-i\Omega'(t_2-t_1)]\}|g\rangle$.

From the wavefunction at $t\rangle t_2+$ we calculate the dipole moment of the atom being followed and label it with the subscript i:

$P_i = \{\exp i\Omega[(t-t_2)-(t_2-t_1)]\} \bullet [1 + \exp -i(\Omega'-\Omega)(t_2-t_1)] + $ c.c..

The dipole density is then obtained by summing over all atoms, yielding

$P(x,t) = \sum_i P_i(t)\delta(x-x_i-v_i t)$ where $\sum_i \rightarrow \int dv g(v)$. The lineshape is given by

$g(v) = (1/\sqrt{2\pi})(1/v_p)\exp -(v-v_s)^2/v_p^2$ where $v_s = (\Delta\Omega/\Omega)c$ is the average velocity of the interacting atoms as determined by the detuning of the laser from the resonance frequency of an atom at rest, the quantity $v_p = (\tau_D/\tau_p)v_D$ is the spread in velocities covered by the laser pulse whose duration is τ_p. The radiated field from those atoms moving at v_i is obtained from

$E \sim \int dx P(x,t+x/c) \sim \sum_i P_i[t+(x_i+v_i t)/(c-v_i)]$.

The relationship between the various parameters is displayed in figure 5 and used to evaluate P_i to order v_i/c. This done we can obtain E and the corresponding integrated intensity which we write as

$\int dt\, I_{echo} \sim 1 + \exp\{(-1/4)[\tau v_p(\Omega-\Omega')/c]^2 \cos[(\Omega-\Omega')\tau(1 + v_s/c)]$

Since the modulation is observed only when both lasers are simultaneously resonant on the same atom the Doppler shift occurs only in the difference frequency. In the limit that the lasers are multimode and the mode density is dense then the response becomes Doppler free and is given by

$\int dt\, I_{echo} \sim 1 + \exp\{-(1/4)[\tau(\Omega'-\Omega)/c]^2(v_D^2+v_p^2/2)\}\cos(\Omega-\Omega')\tau\}$

where v_D is the Doppler velocity. The term involving v_D degrades the echo in Na with a time constant of 400 nsec when the Na temperature is 410 K. This is the dominant relaxation term after lifetime relaxation and only becomes important when the time between the first excitation pulse and the echo exceeds 50 fluorescence lifetimes.

CONCLUSION

It has been demonstrated that incoherent or broadband light is suitable for generating photon echoes and for performing time delayed four wave mixing experiments. The broadband character of the light makes it possible to excite a wide spectrum of resonances. This is useful when relaxation broadening is important. The simplicity of the technique makes it a promising tool for the study of fast relaxation processes and wide band spectroscopy. Insight into the generation of coherent beats with broadband light was obtained by performing a parallel experiment with a pair of narrowband lasers. Doppler free spectroscopy should be possible with this technique.

ACKNOWLEDGMENTS

REFERENCES

1. N. A. Kurnit, I. D. Abella, and S. R. Hartmann, Phys. Rev. Lett. **13**, 567 (1964).
2. R. Beach and S. R. Hartmann, Phys. Rev. Lett. **53**, 663 (1984).
3. Tatsuo Yajima, Norio Morita, and Yuzo Ishida, in XIII International Quantum Electronics Conference Technical Digest, Anaheim, California, 1984 (unpublished), p. 112.
4. S. Asaka. H. Nakatsuka, M. Fujiwara, and M. Matsuoka, Phys. Rev. A **29**, 2286 (1984).
5. R. Beach, D. DeBeer, and S. R. Hartmann, Phys. Rev. A **32**, 3467 (1985)
6. The correponding experiment in Rb. vapor produces a 139 fsec beat. J. E. Golub and T. W. Mossberg, Opt. Lett. **11**, 431(1986)
7. N. Morita and T. Yajima, Phys. Rev. A **30**, 2525 (1984).
8. M. Defour, J. C. Keller, and J. L. LeGouët, JOSA B **3**, 544 (1986)
9. R. Beach, S. R. Hartmann, and R. Friedberg, Phys. Rev. A **25**, 2658 (1982)
10. Masahairo Fujiwara, Ryo Kuroda, and Hiroki Nakatsuka, JOSA B **2** 1634 (1986)
11. D. DeBeer, L. G. Van Wagenen, R. Beach, and S. R. Hartmann, Phys. Rev. Lett. **56**, 1128 (1986).

Comment on "Diffraction-Free Beams"

The recent work of Durnin, Miceli, and Eberly[1] on Diffraction-Free Beams brings to mind the experiment which generates the Poisson Spot.[2] As normally observed by diffracting parallel light around a circular obstacle the Poisson Spot retains its intensity on axis as one recedes from the obstacle while its radius increases linearly. However if the obstacle is placed in the focal plane of a following lens as was the case in ref. 1 the illuminated spot which is observed in the center of the shadow exhibits the general characteristics which are described in Durnin et al. For our experimental conditions, as described below, the spot retains its intensity and sharpness as the observation point is moved over the latter half of the 156 cm. range on which it forms. However, we probably ought not to say that the spot constitutes a diffractionless propagating beam. It is rather a line image, as is shown by the fact that a new obstacle placed in its path does not obliterate the spot further along the axis. In figures 1a and 1b we show the shadow of a .8 cm ball bearing illuminated by a 1.5 cm diameter collimated He-Ne laser light beam as observed on a white card 150 cm from a 25 cm lens which is analogous to the lens in the apparatus of Ref. 1. Figure 1a shows the normal unblocked situation while figure 1b was obtained when a .3 cm diameter rod was inserted accross the beam at a position 40 cm before the white card. The shadow of the rod is clearly visible as is the undisturbed Poisson Spot. This shows that the energy in the spot did not get there by traveling along the axis. The rod only blocks that part of the image lying close enough to the rod so that it cannot "see" the edge of the ball bearing. Further away, light skimming the edge of the ball bearing has a free path to the image position and hence it is not affected.

D. DeBeer and S. R. Hartmann
Columbia Radiation Laboratory and Department of Physics
Columbia University
New York, New York 10027

R. Friedberg
Department of Physics
Barnard College and Columbia University
New York, New York 10027

[1] J. Durnin, J. J. Miceli Jr., and J. H. Eberly, Phys. Rev. Lett. 58, 1499 (1987).
[2] Most books on optics describe diffraction in the shadow of a disk. See, for example, J. B. Marion and M. A. Heald, *Classical Electromagnetic Radiation* (Academic Press, New York, 1980), 2nd ed., p. 377.
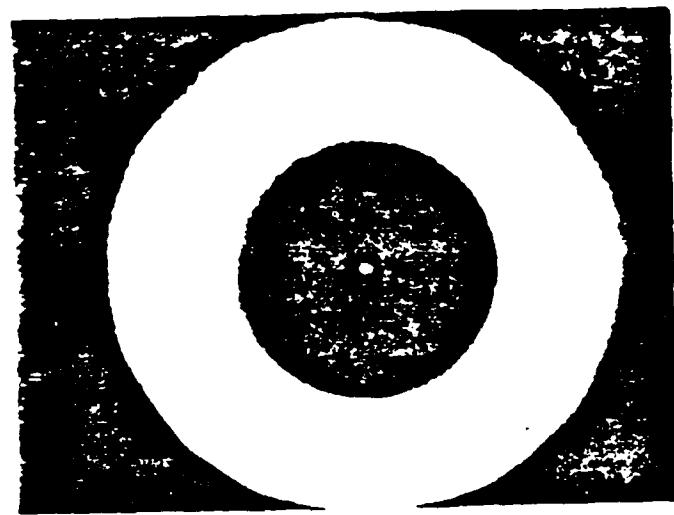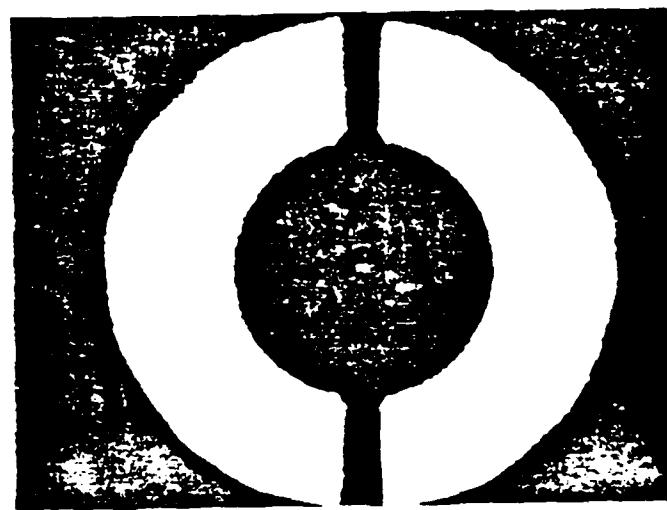
FIG. 1. Photographs showing the bright spot of Poisson in the shadow of a circular object (a) without any additional obstructions; (b) with a rod attempting to block the spot. Both photographs were overexposed to make the Poisson spot more visible.

A DIAGRAMATIC TECHNIQUE FOR CALCULATING RADIATION OF COHERENTLY OR INCOHERENTLY EXCITED TWO-LEVEL ATOMS

R. Friedberg

Department of Physics, Barnard College

New York, New York 10027, U.S.A.

S. R. Hartmann

Columbia Radiation Laboratory, Department of Physics, Columbia University

New York, New York 10027, U.S.A.

New formalism of the Kronig-Penney model with application to superlattices

Hung-Sik Cho and Paul R. Prucnal

Center for Telecommunications Research, Department of Electrical Engineering, Columbia University, New York, New York 10027

(Received 26 March 1987)

A new formalism of the Kronig-Penney model has been developed which is considerably simpler than the conventional one. It gives not only the carrier energy bands but also the wave functions at the edges of each band. Although it is formulated to be applied to superlattices, it is also applicable to bulk materials.

I. INTRODUCTION

The Kronig-Penney model[1] is an idealized method of calculating the energy bands in crystalline solids with periodically varying potentials. It has been applied primarily to bulk semiconductors[2,3] to study the qualitative nature of band structures. Recently, it has also been applied to semiconductor superlattices[4-7] to calculate quantum confined energy subbands. Here, because of conduction- and valence-band discontinuities, smaller–energy-gap layers act as potential wells between larger–energy-gap layers. Though the effective mass of carriers is the same throughout bulk materials, it is different in the well and barrier layers of superlattice structures. This effective mass difference must be taken into account by the appropriate boundary condition at the well/barrier interface. The conventional formalism of the Kronig-Penney model for bulk materials must therefore be modified for superlattices.

Mukherji and Nag[4] considered the effective-mass difference in formulating the Kronig-Penney model, but assumed that the wave function and its first derivative were continuous at the interface. Bastard[5] replaced the assumption of continuity of the first derivative of wave function with an envelope function approximation. He showed that the first derivative of the wave function divided by effective mass, or the probability current, is continuous at the interface under certain conditions. Although several other boundary conditions have been proposed,[8-10] Bastard's is the simplest and most intuitively appealing. The results of particle energy calculations based on Bastard's boundary condition are consistent with the experimental data.[11] Recently, the Kronig-Penney model with Bastard's boundary condition was also used to calculate the band offset of GaAs/Al$_x$Ga$_{1-x}$As superlattices.[12]

The purpose of this paper is to present a new formalism of the Kronig-Penney model developed using Bastard's boundary condition. This formalism is based on the observation that wave functions corresponding to maximum and minimum energies of each band have definite parities. Since this formalism is simpler and easier to manipulate than the conventional formalism, it can be used to find energy bands and wave functions in superlattices and bulk materials by simple numerical analysis. In the following section, both the conventional formalism, and the new formalism of the Kronig-Penney model with Bastard's boundary condition, will be developed and compared. In Sec. III application of the new formalism will be illustrated by finding the energy subbands and envelope wave functions of a GaAs/Al$_x$Ga$_{1-x}$As superlattice. This application will then be discussed.

II. FORMULATION

The periodic potential in a superlattice is not a real atomic potential, but a periodic repetition of different energy gap layers, the thickness of which is usually many times larger than the period of the atomic potential. As a consequence, the actual wave function of a superlattice is a bulk atomic wave function modulated by the envelope wave function of the superlattice potential. The energy bands of a superlattice are composed of a discrete series of subbands induced by the superlattice potential inside a bulk band.

In this section, we will formulate the energy subbands and envelope wave functions induced by the superlattice potential, using a new formalism of the Kronig-Penney model and applying Bastard's boundary condition. Before developing the new formalism, the conventional formalism is first developed for comparison.

A. Conventional formalism

Figure 1 shows the periodic square potential of a superlattice with well thickness a, barrier thickness b, and barrier height V. Particle effective mass is m_a at the well and m_b at the barrier. From the solution of the Schrödinger equation, the envelope wave function in two adjacent periods, $-b < z < a$ and $a < z < d - a$, can be written as

$$\Psi_z = \begin{cases} Ae^{i\alpha(z-a/2)} + Be^{-i\alpha(z-a/2)} \equiv \Psi_a(z), & 0 \le z \le a \\ Ce^{\beta(z+b/2)} + De^{-i\beta(z+b/2)} \equiv \Psi_b(z), & -b \le z \le 0, \end{cases} \tag{1}$$

$$\Psi_z = \begin{cases} e^{ikz}[Ae^{i\alpha(z-a/2-d)} + Be^{-i\alpha(z-a/2-d)}] \equiv \Psi_c(z), & d \le z \le d+a \\ e^{ikz}[Ce^{\beta(z+b/2-d)} + De^{-i\beta(z+b/2-d)}] \equiv \Psi_d(z), & a \le z \le d, \end{cases} \tag{2}$$

where $d = a + b$ is the period of the superlattice, A, B, C, and D are complex numbers, $e^{i\Phi}$ is the phase factor, and z is the axis perpendicular to the superlattice layers. α and β are defined as

$$\alpha \equiv \frac{1}{\hbar}(2m_a E)^{1/2} ,$$

$$\beta \equiv \frac{1}{\hbar}[2m_b(E - V)]^{1/2} ,$$

where E is the particle energy. Ψ_a, Ψ_b, Ψ_c, and Ψ_d are defined only inside a well or a barrier. We note $|\Psi_a(0)| = |\Psi_a(a)|$ and $|\Psi_b(-b)| = |\Psi_b(0)|$ from the symmetry of the wells and barriers, and also note $|\Psi(z + d)| = |\Psi(z)|$ from the periodicity of superlattice potential. Using Bastard's boundary condition at $z = 0$ and $z = a$, i.e.,

$$\Psi_a(0) = \Psi_b(0), \quad \frac{1}{m_a}[\Psi_a'(z)]_{z=0} = \frac{1}{m_b}[\Psi_b'(z)]_{z=0} ,$$

$$\Psi_a(a) = \psi_d(a), \quad \frac{1}{m_a}[\Psi_a'(z)]_{z=a} = \frac{1}{m_b}[\Psi_d'(z)]_{z=a} ,$$

where $\Psi_a'(z)$ and $\Psi_b'(z)$ are the first derivatives of $\Psi_a(z)$ and $\Psi_b(z)$, respectively, and applying the condition that A, B, C, and D should not be zero simultaneously, we obtain

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ \dfrac{\alpha}{m_a} & -\dfrac{\alpha}{m_a} & \dfrac{\beta}{m_a} & -\dfrac{\beta}{m_a} \\ e^{i\alpha a} & e^{-i\alpha a} & e^{i(\Phi - b\beta)} & e^{i(\Phi + b\beta)} \\ \dfrac{\alpha}{m_a}e^{i\alpha a} & -\dfrac{\alpha}{m_a}e^{-i\alpha a} & \dfrac{\beta}{m_b}e^{i(\Phi - b\beta)} & -\dfrac{\beta}{m_b}e^{i(\Phi - b\beta)} \end{vmatrix} = 0 .$$

Note that by Bloch's theorem $\Psi(z) = e^{ikz}u(z)$, where k is the wave number of the envelope function in the z axis and $u(z)$ is a periodic potential satisfying $u(z) = u(z + d)$. It is obvious that $\Psi(z + d) = e^{ikd}e^{ikz}u(z + d)$ is $e^{ikd}\Psi(z)$. Thus Φ is equal to kd. By simplifying the above determinant and replacing Φ by kd, we obtain

$$\cos(kd) = \cos\left[\frac{a}{\hbar}(2m_a E)^{1/2}\right]\cosh\left[\frac{b}{\hbar}[2m_b(V - E)]^{1/2}\right]$$
$$+ \frac{\left[\dfrac{m_a}{m_b}\right]^{1/2}V - \left[\left[\dfrac{m_a}{m_b}\right]^{1/2} + \left[\dfrac{m_b}{m_a}\right]^{1/2}\right]E}{2\sqrt{E(V - E)}}\sin\left[\frac{a}{\hbar}(2m_a E)^{1/2}\right]\sinh\left[\frac{b}{\hbar}[2m_b(V - E)]^{1/2}\right] \quad \text{for } V > E ,$$

$$\tag{3}$$

and

$$\cos(kd) = \cos\left[\frac{a}{\hbar}(2m_a E)^{1/2}\right]\cos\left[\frac{b}{\hbar}[2m_b(E - V)]^{1/2}\right]$$
$$+ \frac{\left[\dfrac{m_a}{m_b}\right]^{1/2}V - \left[\left[\dfrac{m_a}{m_b}\right]^{1/2} - \left[\dfrac{m_b}{m_a}\right]^{1/2}\right]E}{2\sqrt{E(E - V)}}\sin\left[\frac{a}{\hbar}(2m_a E)^{1/2}\right]\sin\left[\frac{b}{\hbar}[2m_b(E - V)]^{1/2}\right] \quad \text{for } V < E , \quad \tag{4}$$

where $\hbar = h/2\pi$ and h is the Planck's constant.

Similar equations have been derived by Bastard,[5] who also included the effect of momentum in the direction parallel to the layers. If continuity of $d\Psi(z)/dz$ is adopted instead of $(1/m)d\Psi(z)/dz$, $(m_a/m_b)^{1/2}$ is replaced by $(m_b/m_a)^{1/2}$ and vice versa, and the overall energy bands shift upward.[11] From Eqs. (3) and (4) the E-k dispersion relations can be readily obtained.

B. New formalism

Even though the conventional formalism yields the energy bands of superlattices or bulk materials, it is quite complicated and cannot yield the envelope wave functions. A new formalism, which will be developed here, is simpler than the conventional formalism, and yields the envelope wave functions corresponding to the edge ener-
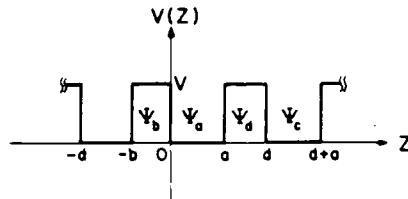
FIG. 1. A schematic view of the periodic square potential of a superlattice. a is the well thickness, b is the barrier thickness, $d = a + b$ is the period of the superlattice, and V is the barrier height. Ψ_a and Ψ_c are the envelope wave functions in the wells at $0 < z < a$ and $d < z < d + a$, respectively. Ψ_b and Ψ_d are the envelope wave functions in the barriers at $-b < z < 0$ and $a < z < d$, respectively.

gies of each band. These envelope wave functions are useful to analyze the superlattice properties.

From Eqs. (3) and (4), the phase difference ($\Phi = kd$) between $\Psi(z)$ and $\Psi(z + d)$ is 0 or $\pm \pi$ for the wave function corresponding to minimum and maximum energy of each band. The phase difference between $\Psi_a(0)$ and $\Psi_c(d)$ is the sum of the phase difference Φ_a between $\Psi_a(0)$ and $\Psi_a(a)$, and the phase difference Φ_b between $\Psi_a(a)$ and $\Psi_d(d)$. When A is different from B or $-B$ in Eq. (1), Φ_a is different from 0 or $\pm \pi$. Similarly, when C is different from D or $-D$ in Eq. (1), Φ_b is also different from 0 or $\pm \pi$. In general, when Φ_a and Φ_b do not have even or odd parity (when $A \neq B$ or $-B$, or $C \neq D$ or $-D$), the total phase difference cannot be 0 or $\pm \pi$. Therefore, the envelope wave functions corresponding to the minimum and maximum energy of each band must have even or odd parity.

1. Odd-index bands ($n = 1, 3, 5, \ldots$)

As will be shown in Fig. 2, the minimum energy of every odd-index band corresponds to $\cos \Phi = 1$ or $\Phi = 0$. Following the above arguments, Φ_a and Φ_b should both be even or both be odd at minimum energies. Note, how-
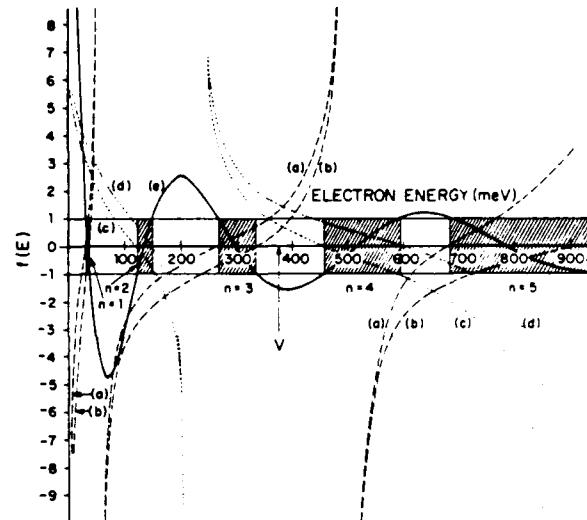

FIG. 2. The first five electron energy subbands of a GaAs/Al$_x$Ga$_{1-x}$As superlattice with $x = 0.5$, $a = 100$ Å, and $b = 25$ Å. $V = 375$ meV is the barrier potential height. Dashed curves (a) and (b) correspond to the functions in the left-hand side of Eqs. (7) and (8), and (11) and (12), respectively. Dotted curves (c) and (d) correspond to the functions in the left-hand side of Eqs. (15) and (16), and (19) and (20), respectively. Solid curve (e) corresponds to the right-hand side of Eqs. (3) and (4), respectively. n is the band index.

ever, that at large b each band in a well degenerates to an energy level with vanishing bandwidth. Thus Ψ_a^{min} (corresponding to minimum energy) and Ψ_a^{max} (corresponding to maximum energy) become identical. Hence, the parities of Ψ_a^{min} and Ψ_a^{max} must be the same for any value of b. Furthermore, Ψ_a is even for the first (ground) state, odd for the second state, and alternating even and odd with increasing n. Therefore, both Ψ_a^{min} and Ψ_b^{min} of every odd-index band have even parity, and can be written as

$$\Psi_a^{min}(z) = A \cos \left| \alpha \left| z - \frac{a}{2} \right| \right|, \quad \Psi_b^{min}(z) = B \cosh \left| \delta \left| z - \frac{b}{2} \right| \right| \quad \text{for } V > E \ . \tag{5}$$

$$\Psi_a^{min}(z) = A \cos \left| \alpha \left| z - \frac{a}{2} \right| \right|, \quad \Psi_b^{min}(z) = B \cos \left| \beta \left| z - \frac{b}{2} \right| \right| \quad \text{for } V < E \ . \tag{6}$$

where $\delta = i\beta = [2m_b(V - E)]^{1/2}/\hbar$.

Using Bastard's boundary condition at $z = 0$, i.e.,

$$\Psi_a(0) = \Psi_b(0)$$

and

$$\frac{1}{m_a} [\Psi_a'(z)]_{z=0} = \frac{1}{m_b} [\Psi_b'(z)]_{z=0} \ ,$$

we obtain

$$A\cos\left[\frac{a\alpha}{2}\right]=B\cosh\left[\frac{b\delta}{2}\right], \quad \frac{A}{m_a}\alpha\sin\left[\frac{a\alpha}{2}\right]=\frac{B}{m_b}\delta\sinh\left[\frac{b\delta}{2}\right] \quad \text{for } V>E ,$$

$$A\cosh\left[\frac{a\alpha}{2}\right]=B\cos\left[\frac{b\delta}{2}\right], \quad \frac{A}{m_a}\alpha\sin\left[\frac{a\alpha}{2}\right]=\frac{-B}{m_b}\beta\sin\left[\frac{b\beta}{2}\right] \quad \text{for } V<E ,$$

which in turn gives the minimum energy of every odd-index band:

$$\tan\left[\frac{a}{2\hbar}(2m_a E_{\min})^{1/2}\right]-\left[\frac{m_a}{m_b}\left[\frac{V}{E_{\min}}-1\right]\right]^{1/2}\tanh\left[\frac{b}{2\hbar}[2m_b(V-E_{\min})]^{1/2}\right]=0 \quad \text{for } V>E , \tag{7}$$

$$\tan\left[\frac{a}{2\hbar}(2m_a E_{\min})^{1/2}\right]+\left[\frac{m_a}{m_b}\left[1-\frac{V}{E_{\min}}\right]\right]^{1/2}\tan\left[\frac{b}{2\hbar}[2m_b(E_{\min}-V)]^{1/2}\right]=0 \quad \text{for } V<E . \tag{8}$$

The smallest solution of Eq. (7) corresponds to $n=1$ (ground subband), the next solution to $n=3$, etc. The band index corresponding to the smallest solution of Eq. (8) depends on how many bands are confined inside the well.

To find the maximum energy of each odd-index band, we note that $\cos\Phi$ is -1 or Φ is $\pm\pi$. Φ_a is 0 because Ψ_a is even. Since $\Phi_a+\Phi_b$ is $\mp\pi$, Φ_b is $\mp\pi$, or Φ_b is odd. Hence, the wave functions can be written as

$$\Psi_a^{\max}(z)=A\cos\left[\alpha\left[z-\frac{a}{2}\right]\right], \quad \Psi_b^{\max}=B\sinh\left[\delta\left[z+\frac{b}{2}\right]\right] \quad \text{for } V>E , \tag{9}$$

$$\Psi_a^{\max}(z)=A\cos\left[\alpha\left[z-\frac{a}{2}\right]\right], \quad \Psi_b^{\max}=B\sin\left[\beta\left[z+\frac{b}{2}\right]\right] \quad \text{for } V<E . \tag{10}$$

Following the same procedure as above, we obtain the maximum energy of every odd-index band:

$$\tan\left[\frac{a}{2\hbar}(2m_a E_{\max})^{1/2}\right]-\left[\frac{m_a}{m_b}\left[\frac{V}{E_{\max}}-1\right]\right]^{1/2}\coth\left[\frac{b}{2\hbar}[2m_b(V-E_{\max})]^{1/2}\right]=0 \quad \text{for } V>E , \tag{11}$$

$$\tan\left[\frac{a}{2\hbar}(2m_a E_{\max})^{1/2}\right]-\left[\frac{m_a}{m_b}\left[1-\frac{V}{E_{\max}}\right]\right]^{1/2}\cot\left[\frac{b}{2\hbar}[2m_b(E_{\max}-V)]^{1/2}\right]=0 \quad \text{for } V<E . \tag{12}$$

2. Even-index bands (n = 2,4,6,...)

For even-index bands, the minimum energy corresponds to $\cos\Phi=-1$ or $\Phi=\pm\pi$. Since Ψ_a is odd and Φ_a is $\mp\pi$, $\Phi_b=\Phi-\Phi_a$ is 0. Thus Ψ_b is even, and we can write

$$\Psi_a^{\min}(z)=A\sin\left[\alpha\left[z-\frac{a}{2}\right]\right], \quad \Psi_b^{\min}=B\cosh\left[\delta\left[z+\frac{b}{2}\right]\right] \quad \text{for } V>E , \tag{13}$$

$$\Psi_a^{\min}(z)=A\sin\left[\alpha\left[z-\frac{a}{2}\right]\right], \quad \Psi_b^{\min}=B\cos\left[\beta\left[z+\frac{b}{2}\right]\right] \quad \text{for } V<E . \tag{14}$$

Using the boundary condition and simplifying as above, we obtain the minimum energy of every even-index band:

$$\cot\left[\frac{a}{2\hbar}(2m_a E_{\min})^{1/2}\right]+\left[\frac{m_a}{m_b}\left[\frac{V}{E_{\min}}-1\right]\right]^{1/2}\tanh\left[\frac{b}{2\hbar}[2m_b(V-E_{\min})]^{1/2}\right]=0 \quad \text{for } V>E , \tag{15}$$

$$\cot\left[\frac{a}{2\hbar}(2m_a E_{\min})^{1/2}\right]-\left[\frac{m_a}{m_b}\left[1-\frac{V}{E_{\min}}\right]\right]^{1/2}\tan\left[\frac{b}{2\hbar}[2m_b(E_{\min}-V)]^{1/2}\right]=0 \quad \text{for } V<E . \tag{16}$$

The smallest solution of Eq. (15) corresponds to $n=2$, the next to $n=4$, etc. The band index corresponding to the smallest solution of Eq. (16) depends on how many bands are confined inside the well.

To find the maximum energy of every even-index band we note that $\cos\Phi$ is 1 or Φ is 0. Since Ψ_a is odd and Φ_a is $\pm\pi$, $\Phi_b=\Phi-\Phi_a$ is $\pm\pi$. Thus Ψ_b is odd. Hence, we can write the wave functions as

$$\Psi_a^{\max}(z)=A\sin\left[\alpha\left[z-\frac{a}{2}\right]\right], \quad \Psi_b^{\max}=B\sinh\left[\delta\left[z+\frac{b}{2}\right]\right] \quad \text{for } V>E , \tag{17}$$

$$\Psi_a^{\max}(z)=A\sin\left[\alpha\left[z-\frac{a}{2}\right]\right], \quad \Psi_b^{\max}=B\sin\left[\beta\left[z+\frac{b}{2}\right]\right] \quad \text{for } V<E . \tag{18}$$

Following the same procedure as above, we obtain the maximum energy of every even-index band:

$$\cot\left[\frac{a}{2\hbar}(2m_a E_{max})^{1/2}\right]+\left[\frac{m_a}{m_b}\left[\frac{V}{E_{max}}-1\right]\right]^{1/2}\coth\left[\frac{b}{2\hbar}[2m_b(V-E_{max})]^{1/2}\right]=0 \quad \text{for } V>E , \tag{19}$$

$$\cot\left[\frac{a}{2\hbar}(2m_a E_{max})^{1/2}\right]+\left[\frac{m_a}{m_b}\left[1-\frac{V}{E_{max}}\right]\right]^{1/2}\cot\left[\frac{b}{2\hbar}[2m_b(E_{max}-V)]^{1/2}\right]=0 \quad \text{for } V<E . \tag{20}$$

Even though formulas similar to Eqs. (7) and (19) have been previously derived by Masselink et al.,[11] their relationship to the Kronig-Penney model has not been clarified. In addition, they are not a complete formalism because they cannot yield the maximum energies of odd-index bands and minimum energies of even-index bands, and the corresponding envelope wave functions.

The question might arise why only one boundary condition at $z=0$ was used in developing the new formalism, whereas boundary conditions at both $z=0$ and $z=a$ were used to derive Eqs. (3) and (4). The explanation lies in the main difference between the two formalisms. Since the conventional formalism is concerned with all the energies from 0 to ∞, Eqs. (3) and (4) indicate that Φ ($=kd$) can be any real value from 0 to 2π (for energies inside bands) or any complex value (for energies outside bands). Thus there are no definite parities of Ψ_a and Ψ_b and no symmetry relationship between $\Psi_a(0)$ and $\Psi_a(a)$ or $\Psi_b(-b)$ and $\Psi_b(0)$. Therefore, in Fig. 1, the boundary condition at $z=0$ and $z=a$ are different and both boundary conditions must be used. On the other hand, the new formalism is only concerned with the band-edge energies, at which Φ can be only 0 or $\pm\pi$. Thus Ψ_a and Ψ_b have definite parities. Since $\Psi_a(0)=\pm\Psi_a(a)$ and $\Psi_b(0)=\pm\Psi_d(a)$, the boundary condition at $z=0$, which determines the relationship between Ψ_a and Ψ_b, is identical to that at $z=a$. This is the reason we need only one boundary condition in developing the new formalism.

When b becomes sufficiently large for $V>E$, Eqs. (7) and (11) degenerate to

$$\tan\left[\frac{a}{2\hbar}(2m_a E)^{1/2}\right]-\left[\frac{m_a}{m_b}\left[\frac{V}{E}-1\right]\right]^{1/2}=0$$

$$\text{for } n=1,3,5,\dots$$

and Eqs. (15) and (19) to

$$\cot\left[\frac{a}{2\hbar}(2m_a E)^{1/2}\right]+\left[\frac{m_a}{m_b}\left[\frac{V}{E}-1\right]\right]^{1/2}=0$$

$$\text{for } n=2,4,6,\dots .$$

These results are, of course, identical to the single-well results.[13]

III. APPLICATION AND DISCUSSION

As an illustration of the application of the new formalism, we have calculated the electron energy subbands of a GaAs/Al$_x$Ga$_{1-x}$As superlattice with $x=0.5$, $a=100$ Å, and $b=25$ Å by using both the conventional formalism [Eqs. (3) and (4)] and the new formalism [Eqs. (5) to (20)].

We have also calculated the envelope wave functions using the new formalism.

Figure 2 shows the first five electron energy subbands. We have taken $\Delta E_g=0.476+0.125x+0.143x^2$ eV for $0.45<x<1$,[14] as the band-gap difference between GaAs and Al$_x$Ga$_{1-x}$As at room temperature; $m_a=0.067m_0$ and $m_b=(0.067+0.083x)m_0$,[15] where m_0 is the free electron mass, as the electron effective mass at well and barrier layers. We have assumed 65%:35% band offset ratio[11,16,17] between the conduction and valence band. Thus the barrier potential V is given as $0.65\Delta E_g=375$ meV. As shown in Fig. 2, the first three bands are confined inside the well and the remaining two bands are not. Curve (a) in Fig. 2 corresponds to the function in the left-hand side of Eqs. (7) and (8) for $E<V$ and $E>V$, respectively. This curve meets the E axis at 31, 268, and 684 meV. Since Eqs. (7) and (8) give the minimum energies of odd-index bands, these three values are the minimum energy of the first, third, and fifth subband, respectively. Curve (b) corresponds to the functions in the left-hand side of Eqs. (11) and (12). Obviously, it gives the maximum energy of the first, third, and fifth subband. Similarly, curves (c) and (d) correspond to the functions in the left-hand side of Eqs. (15) and (16), and (19) and (20), respectively.
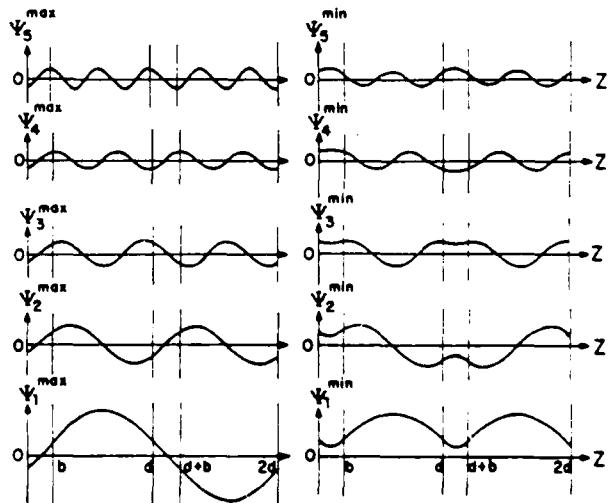
FIG. 3. Envelope wave functions corresponding to the edge energies of the first five subbands shown in Fig. 2. Ψ^{max} is the wave function corresponding to the maximum edge of the first subband, and Ψ^{min} is that corresponding to the minimum edge of the first subband. etc. $b=25$ Å is the barrier thickness; $a=d-b=100$ Å is the well thickness. The amplitude of the wave functions has arbitrary units.

They give the minimum and maximum energies of the second and fourth subband, respectively. All these curves are obtained from the new formalism. Curve (e) is obtained from the conventional formalism. It corresponds to the functions in the right-hand side of Eqs. (3) and (4) for $E < V$ and $E > V$, respectively. We can easily identify the first five subbands.

If we define $f(E)$ as the longitudinal axis of the curves $(a)-(e)$, we note that curve (e) meets the $f(E) = 1$ or -1 line at exactly the same points on the E axis as the minimum and maximum energies found from curves (a), (b), (c), and (d). This indicates that the energy bands found from the new formalism are identical to those found from the conventional formalism, as they must be. Even though we have plotted complete curves of functions in Eqs. (7), (8), (11), (12), (15), (16), (19), and (20) to find the subbands by the new formalism, usually it is not necessary. The energy bands can be easily found by simple numerical or graphical methods. Of course this formalism can also be used for bulk materials with $m_a = m_b$.

In Fig. 3 the envelope wave functions are presented, which correspond to the minimum and maximum energies of the first five subbands of the superlattice in Fig. 2. These wave functions have been found by using the minimum and maximum energy of each band in calculating α, β, and δ; also the boundary condition $\Psi_a(0) = \Psi_b(0)$ is applied to Eqs. (5), (6), (9), (10), (13), (14), (17), and (18), thus eliminating A or B. Although these wave functions are not normalized, $P_b = I_b/(I_a + I_b)$, where $I_a = \int_0^a |\Psi_a|^2 dz$ and I_b

$= \int_{-b}^0 |\Psi_b|^2 dz$, gives the probability to find an electron at a barrier layer. The calculated values of P_b for this superlattice have been found to be 5.7%, 12%, 25%, 38%, and 43% for minimum energy of the first five bands, and 1.3%, 5.3%, 8%, 13%, and 18% for maximum energy of the first five bands. They are an increasing function of n, because electrons become more energetic with increasing n and are more probable to leak to the barriers. P_b^{min} for Ψ_b^{min} is larger than P_b^{max} for Ψ_b^{max} of the same band, because Ψ_b^{min} has even parity and Ψ_b^{max} has odd parity.

We observe another general property of the envelope wave functions in wells and barriers of superlattices: When $E < V$, wave functions at the wells, corresponding to both E_{min} and E_{max} of the nth subband, have $n - 1$ zeros (meet z axis $n - 1$ times). Those at the barriers have 0 (for E_{max}) or 1 (for E_{min}) zero. When $E > V$, wave functions at the wells still have $n - 1$ zeros; those at the barriers have a different number of zeros for different b, ranging from 0 for small b, to infinity for infinitely large b. In Fig. 3, however, we see only 0 or 1 zero inside the barrier, because the barrier is sufficiently thin to have only 0 or 1 zero.

In conclusion, we can calculate both the energy bands and the wave functions corresponding to the edge energies of each band, of superlattices or bulk materials, using the new formalism of the Kronig-Penney model. Moreover, the calculation of energy bands can be done more easily by the new formalism than by the conventional formalism. The new formalism is very useful for the analysis of superlattices and application to device designs.

[1]R. de L. Kronig and W. J. Penney, Proc. R. Soc. London, Ser. A **130**, 499 (1930).

[2]J. P. McKelvey, *Solid State and Semiconductor Physics* (Harper and Row, New York, 1966).

[3]K. Seeger, *Semiconductor Physics*, 3rd ed. (Springer-Verlag, Berlin, 1985).

[4]D. Mukherji and B. R. Nag, Phys. Rev. B **12**, 4338 (1975).

[5]G. Bastard, Phys. Rev. B **24**, 5693 (1981).

[6]J. N. Schulman and Y. C. Chang, Phys. Rev. B **24**, 4445 (1981).

[7]B. A. Vojak, W. D. Laidig, N. Holonyak, Jr., M. D. Camras, J. J. Coleman, and P. D. Dapkus. J. Appl. Phys. **52**, 621 (1981).

[8]T. Ando and S. Mori, Surf. Sci. **113**, 124 (1982).

[9]S. R. White and L. J. Sham, Phys. Rev. Lett. **47**, 879 (1981).

[10]Q. Zhu and H. Kroemer, Phys. Rev. B **27**, 3519 (1983).

[11]W. T. Masselink, P. J. Pearah, J. Klem, C. K. Peng, H. Morkoç, G. D. Sanders, and Y. C. Chang, Phys. Rev. B **32**, 8027 (1985).

[12]A. Chomette, B. Deveaud, M. Baudet, P. Auvray, and A. Regreny, J. Appl. Phys. **59**, 3835 (1986).

[13]H. Kawai, K. Kaneko, and N. Watanabe, J. Appl. Phys. **56**, 463 (1984).

[14]H. C. Casey, Jr. and M. B. Panish, *Heterostructure Lasers* (Academic, New York, 1978), Pt. II.

[15]S. Adachi, J. Appl. Phys. **58**, R1 (1985).

[16]D. Arnold, A. Ketterson, T. Henderson, J. Klem, and H. Morkoç, Appl. Phys. Lett. **45**, 1237 (1984).

[17]H. Okumura, S. Misawa, S. Yoshida, and S. Gonda, Appl. Phys. Lett. **46**, 377 (1985).

THE ELECTROCHEMICAL SOCIETY
CAMERA READY COPY USED FOR

LASER-ASSISTED PLASMA ETCHING

W. Holber*, J.O. Chu*, D. Gaines, A. Nahata,
and R.M Osgood, Jr.

Microelectronics Sciences Laboratories
and Columbia Radiation Laboratory
New York, New York 10027

*Presently at
IBM T.J. Watson Research Center,
Yorktown Heights, New York 10598

ABSTRACT

Laser radiation has been used to enhance
both gas- and surface-phase phenomena
occurring in a plasma reactor.   As a
result,   resistless   patterning   and
light-enhanced plasma etch rates have
been demonstrated.

INTRODUCTION

In the plasma etching processes now routinely used to
fabricate semiconductor devices, a resist mask is used to
define an area to be etched, and a glow discharge
generates reactive species which etch the exposed area.
Depending on the myriad process parameters, such as gas
components, gas pressure, and RF discharge frequency, the
etch can have varying amounts of selectivity, anisotropy,
and damage.   However, a trade-off among these factors
often exists, making it difficult to separately optimize
all of them.   For example, while increasing the ion flux
to the substrate can increase the degree of anisotropy,
it can also decrease the etch material selectivity and
increase the amount of ion damage.   In addition, the
requirement of a resist mask to define the etch region
necessitates many extra steps in the processing sequence.

Recent experiments have shown that light-assisted
chemistry can also be used to accomplish dry etching. In
addition, the use of light to alter the properties of
electric discharges, viz, in gas lasers, and to effect
surface reactions has been previously demonstrated. As a
result, it is reasonable to expect that the introduction
of an external light source to a plasma reactor may
provide a useful tool for improving the performance of a
plasma process.

In a set of experiments which will be described here, we have used laser light to control the etching rate of silicon and gallium arsenide in a plasma etching environment. Various light-assisted processes have been examined, including those relying on gas- and surface-phase interactions. Both substantial etch-rate enhancements and micrometer-scale maskless patterning have been observed.

CW LASER-ENHANCED PLASMA ETCHING OF SILICON

In the study of laser enhanced plasma etching of silicon, cw laser light of various wavelengths and intensities was weakly focused onto the surface of a single-crystal silicon sample, doped to varying degrees. The experiments are described in detail elsewhere (1,2). Briefly, it was found that at high laser power densities (greater than 10 kW/cm^2), there is an etch enhancement, which appears to be primarily thermal in nature, and is independent of incident wavelength, doping type, and doping level. At low laser power densities (less than 10 kW/cm^2) however, the effect appears to be predominately a photochemical one, since the temperature rise due to the laser is quite low (less than 20 $^\circ$C), and the effect is now dependent on the wavelength of the light and the doping level of the sample.

The experimental results are consistent with a process that is controlled by the flow of photogenerated carriers to the surface. Figure 1 demonstrates the photochemical nature of the etch rate enhancement at low laser power densities. Here it can be seen that for heavily doped silicon, either n- or p-type, the photoinduced-etch enhancement is greater for the near-uv laser light than for the visible laser light. Within the experimental accuracy of this work (about 20%), the results are the same for either n- or p-type silicon. The etch rate without laser enhancement for the heavily doped n- and p-type samples is also approximately the same for the operating conditions used in this particular reactor.

This behavior is in marked contrast with the results of another group, which studied purely photon-enhanced etching of silicon by chlorine atoms (3). In that experiment, it was found that the dark etch rate is much greater for n-type than p-type silicon, but that the

relative photon-enhancement is much greater for the p-type than for the n-type material.

Although the precise mechanism by which the photogenerated carriers enhance the etching of silicon in the CF_4/O_2 plasma has not yet been determined, there are a number of possibilities. These include increasing the reaction rate between Si and SiF_x groups (as pointed out by Houle (4) in another study of purely light- induced etching), increasing the rate of subsurface fluorination of the silicon, and removal of adsorbed layers, such as hydrocarbons, on the silicon surface, which may limit the reaction rate by blocking surface sites. Note that all three of these mechanisms result in increased surface desorption in the illuminated region.

In order to examine whether a more precisely defined surface layer could be affected by laser light, polymer deposition was induced by adding hydrogen to CF_4 in the same etching reactor. It was found that both the plasma etch rate and the etch-rate enhancement due to the laser were altered.

First, Figure 2 shows the reduction of the dark etch rate with the addition of hydrogen. It is thought that this reduction in etch rate is due to the onset of two successive processes; H atoms scavenging F atoms from the system via the formation of HF, a nonreactive gas for polymer etching, and polymer deposition on the surface due to the altered gas-phase chemistry (5,6), respectively.

Second, when the laser is used to illuminate the surface of the silicon during the etch process, a clearly defined enhancement is again observed. Figure 3 shows the etch-rate enhancement due to the laser. The most prominent feature is the sharp rise in enhancement at 15-mTorr H_2 partial pressure. According to a previous study, the F concentration in the gas will begin to rise after its initial fall at about this hydrogen concentration (6). As a result, the etch rate will then be limited only by polymer deposition. Under these conditions, the laser light incident to the surface will enhance the etch rate due to the desorption of polymeric material. The reduction in polymer deposition due to the laser light can be seen visually by a color change in the polymer layer in the vicinity of the laser spot, and through Auger electron spectroscopy, which shows a marked reduction in polymer coverage in the illuminated region.

An important application of this technique is that it can be used to accomplish resistless processing. Since the strongly etched region is defined by the laser light, projection of a masked beam on the surface directly patterns the etched region. This form of in situ processing was demonstrated by patterning the laser beam with a simple wire grid. The resolution, which was limited by the demagnification of the projection lens, was several micrometers.

The role of the laser light in preventing polymer film growth appears to again be due to photogenerated carriers in the silicon. The low laser power density used ($\sim$10 W/cm^2) would appear to preclude a purely thermal effect. Also, in the same reactor, the laser light does not have any effect on deposition on sapphire, an insulator which does not absorb light at this wavelength, but would be subject to any surface photochemical effects in the same reactor.

In a second set of experiments, we examined the effect of laser light on controlling the plasma discharge, either through its chemical species, or through its electrical properties. These studies involved laser-enhanced plasma etching of GaAs. The output from a 193-nm excimer laser is used to illuminate the discharge region between two electrodes in a pulsed d.c. discharge. Both the laser and plasma are pulsed at a repetition rate of 30 Hz. The laser power is typically 20 mJ/pulse, and the discharge is typically operated at 600V and a pulse width of 200 usec. The gas used is HBr, at a pressure of 5 torr and a flow rate of 25 sccm.

Previous work describes some aspects of the experimental arrangement in more detail and discusses the etch results obtained with laser illumination alone, i.e., without a plasma discharge (7). A recent extension of these purely laser studies has shown that partially brominated surface layers play an important role in the etching process. In particular, when the etched samples are examined using x-ray photoelectron spectroscopy (XPS), it is found that the chemical composition and the depth of the surface layer varies according to the processing conditions employed. Figure 4 shows the arsenic, gallium and bromine peaks, obtained by etching under several different substrate temperatures. The exact energy of the arsenic and gallium peak depends on the local bonding environment. For example, gallium from

the substrate (GaAs) can be distinguished from gallium
due to $GaBr_3$. However, the peaks due to $GaBr_3$ and
Ga_2O_3 can not be resolved in this apparatus. The
same is true for the arsenic peaks. These results are
consistent with what has been found by other researchers
in different etching systems. For example, in fluorine
etching of silicon, it has been found that subsurface
fluorination occurs, also to a depth of about 20 or
30 (8).

Adding the laser illumination to the discharge region
has several effects on the etching, particularly as a
result of phenomena involving the adsorbed layers.
First, the etch rate is greatly increased over the plasma
etch rate, especially at temperatures high enough so that
the process is not limited by desorption of the etch
reaction products, namely $GaBr_3$ and $AsBr_3$. This is
shown in Figure 5, where it can be seen that the laser
enhances the plasma etch rate by almost an order of
magnitude at $100^\circ C$. Note that in this case, the laser
beam is aimed over the substrate in the pulsed plasma
discharge. Calculations indicate that at the lower
temperatures, where the laser has little effect on the
etching, as shown in Figure 5, the etch rate is limited
by desorption of the $GaBr_3$ products from the GaAs
surface, so the creation of more gas-phase bromine by the
laser has little effect on the etch rate. However, at
the higher temperatures, where desorption is not as
limiting a process, the laser has a large effect, since
the bromine it creates can be used effectively in the
etching.

Second, etching with the plasma alone or the laser
alone yields etch profiles which are quite different.
The etch profile obtained using just the laser follows
crystallographic lines, while that obtained using just
the plasma is more nearly vertical. With the laser and
plasma together, an intermediate etch profile is obtained
which can be more or less anisotropic depending on the
relative contribution of the laser versus that of the
plasma. This behavior is due to the laser changing the
density of neutral bromine atoms in the plasma region,
relative to the density of ions created by the
discharge. While the neutral species tend to cause
crystallographic etching, the ions tend to cause etching
which is more directional.

Laser light can also be used to control the
electrical properties of the discharge. Ultraviolet
controlled or sustained discharges have been used to

obtain abnormally low ratios of electrical field to
neutral particle density. This low ratio can, for
example, yield a more efficient discharge for obtaining
population inversion in a laser. In this connection, we
have also begun to investigate the use of UV excimer
laser light to influence the electrical parameters in the
pulsed, HBr discharge described in the paragraph above.

In particular, preliminary work has been done in
studying changes in plasma characteristics due to the
laser illumination. The technique used was to irradiate
the discharge region with the laser, and at some fixed,
time delay apply a voltage pulse to the discharge
plates. The voltage applied is below the threshold at
which the gas will break down without any laser
illumination. Care was taken to ensure that the laser
light did not hit the electrodes.

The first experiments showed that 193-nm light was
effective in lowering the breakdown voltage in the
plasma. Thus, if a pulse of excimer light preceded the
application of the pulsed waveform to the discharge
plate, the breakdown of the voltage could be lowered as
much as 50%. Further, the magnitude of the reduction in
voltage depended on the time between laser pulse and the
discharge, and the gas conditions in the discharge
region. For example, the higher the HBr pressure, the
longer the duration of the interval of low
discharge-breakdown voltage. The details of this
behavior are in accord with a diffusion-limited loss of
the photogenerated species. In addition, the
conductivity or carrier density in the plasma is found to
be directly proportional to the laser power, when the
193-nm laser pulse and the application of the voltage
pulse to the discharge plates are made to coincide. This
behavior is shown in Figure 6, which displays the
variation in discharge current with laser power, for a
fixed-voltage pulse. Both of these experiments
illustrate that UV light can lower the operating voltage
within the plasma, and yet still obtain significant
discharge current.

In summary, we have shown that laser light can be
used to control surface desorption, chemical species, and
electrical parameters of a plasma reactor. These laser
control phenomena have been used to obtain the practical
advantages of maskless patterning and higher etch rate
(at reduced driving voltage) in a plasma etching
environment.

ACKNOWLEDGMENTS

REFERENCES

1. W. Holber, G. Reksten, and R.M. Osgood, Jr., Appl. Phys. Lett. 46, 201 (1985).

2. Grace M. Reksten, W. Holber, and R.M. Osgood, Jr., Appl. Phys. Lett. 48, 551 (1986).

3. Y. Horiike, M. Sekine, K. Horioka, T. Arikado, and H. Okano, "Photo-excited Dry Etching for VLSI's," in Extended Abstracts of the 16th International Conference on Solid-State Devices and Materials, Kobe, (1984).

4. F.A. Houle, J. Chem. Phys. 10, 4851 (1984).

5. D.L. Flamm, V. M. Donnelly, and D E. Ibbotson, J. Vac. Sci. Technol. B. 1, 23 (1983).

6. Riccardo d'Agostino, J. Vac. Sci. Technol. A 3, 2627 (1985).

7. P.D. Brewer, D. McClure, and R.M. Osgood, Jr., Appl. Phys. Lett. 47, 310 (1985).

8. F.R. McFeeley, J.F. Morar, and F.J. Hempsel, Surface Science 165 (1986).
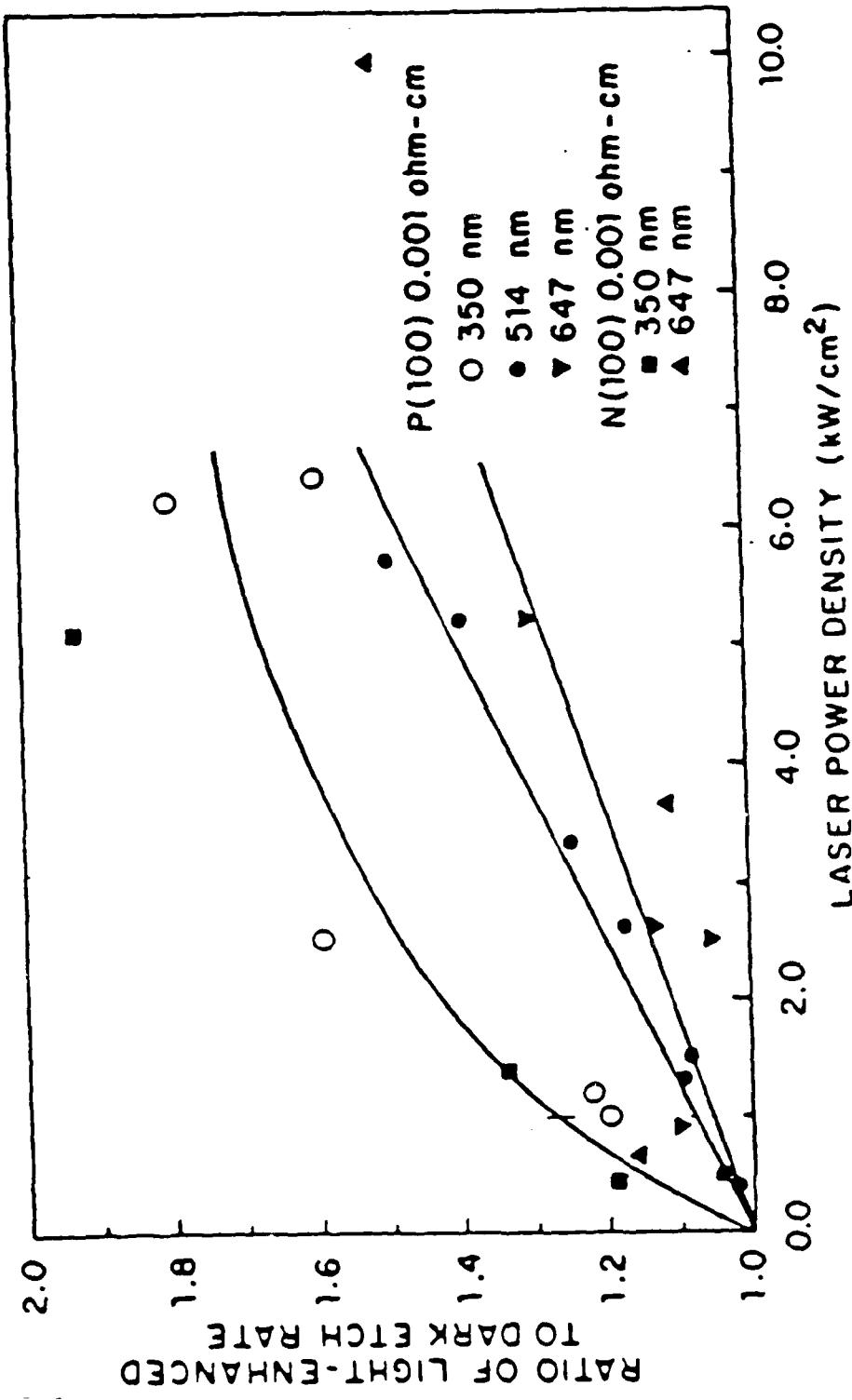
Figure 1: Normalized photoinduced etch rate enhancement vs laser power density. The gas mixture used is 120 mTorr of CF_4/O_2 and the plasma, RF power density is approximately 0.1 W/cm^2.
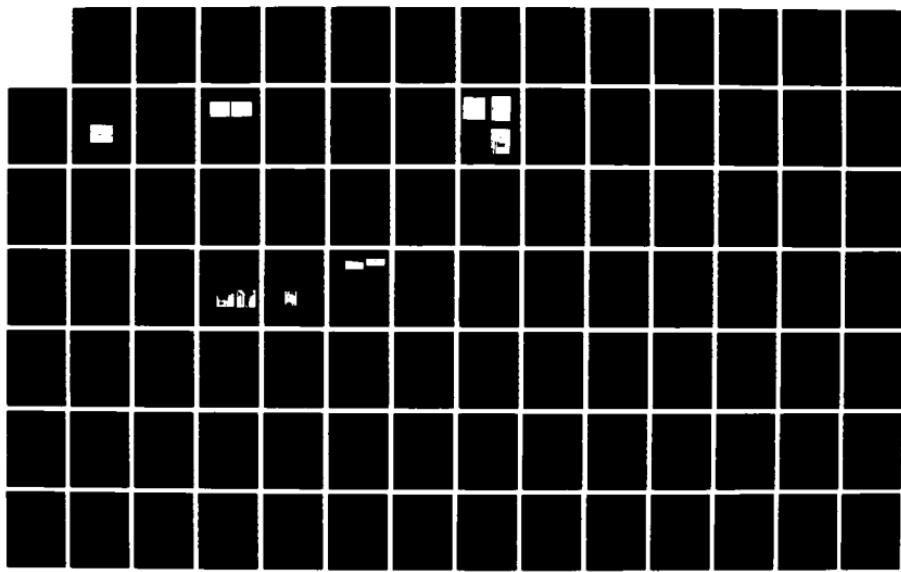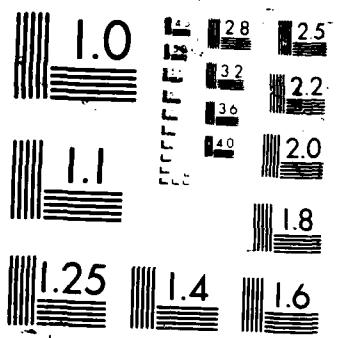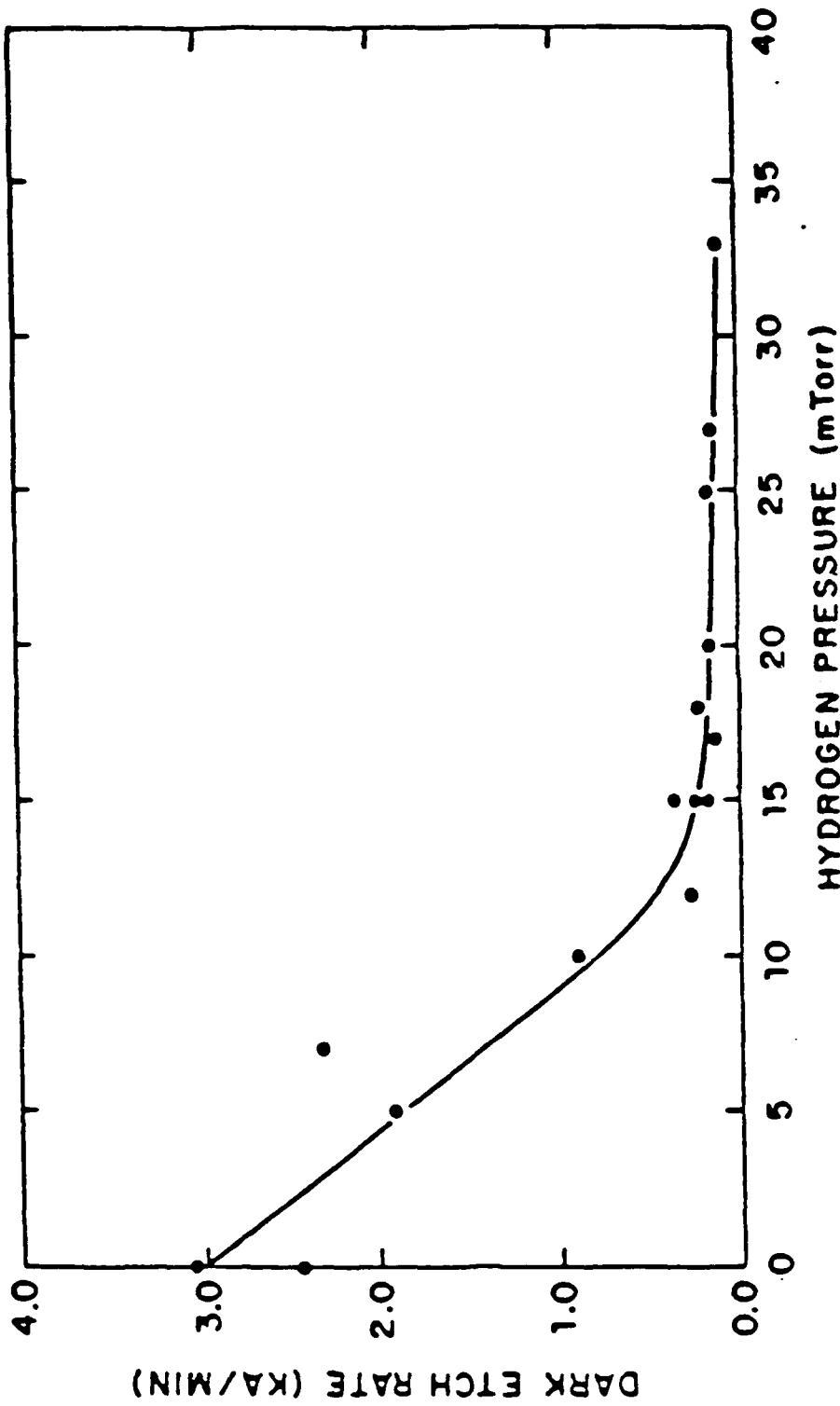
Figure 2: Plasma etch rate as a function of added hydrogen gas. The base gas is CF_4 at 40 mTorr and the RF power density is approximately 0.05 W/cm.

Figure 3: Normalized photoinduced etch rate enhancement vs added hydrogen pressure. The base gas is CF_4 at 40 mTorr and the RF power density is approximately 0.05 W/cm^2. The laser wavelength is 350 nm, and the laser power density is approximately 10 W/cm^2.
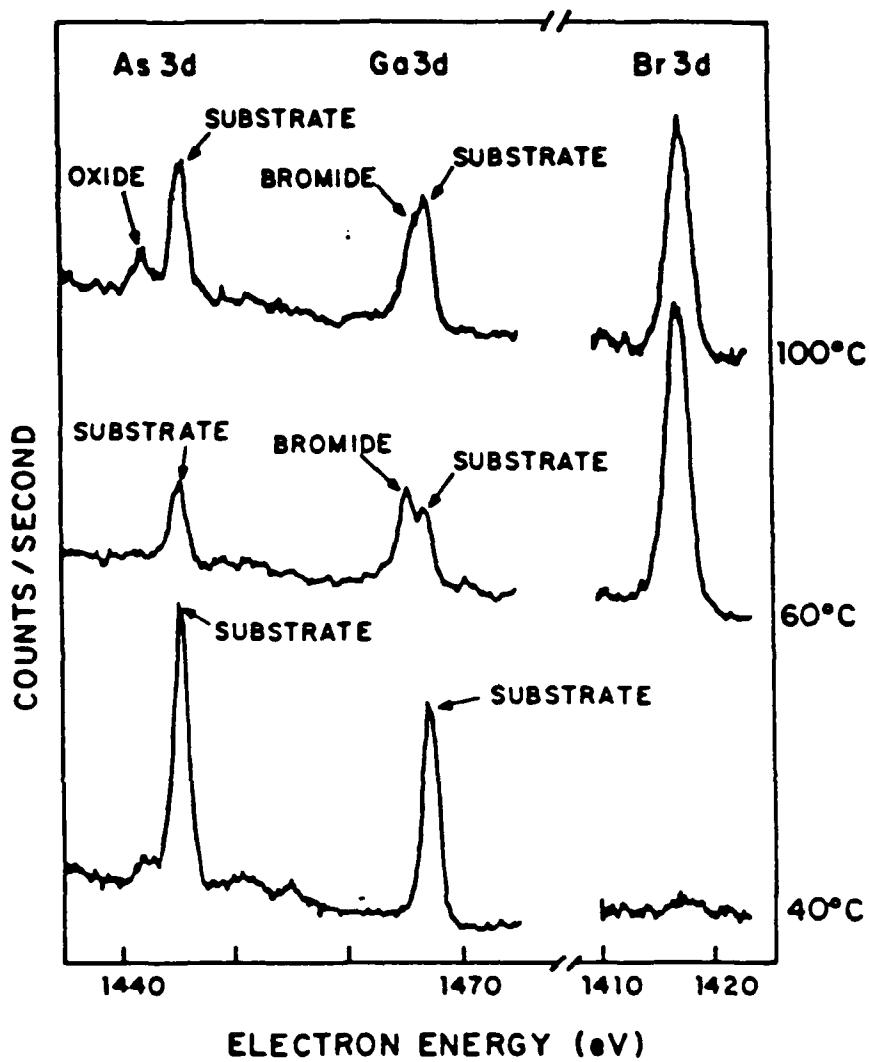
Figure 4: XPS spectra of etched GaAs surface, with
varying substrate temperature during the
etch process. The laser orientation is
indirect.
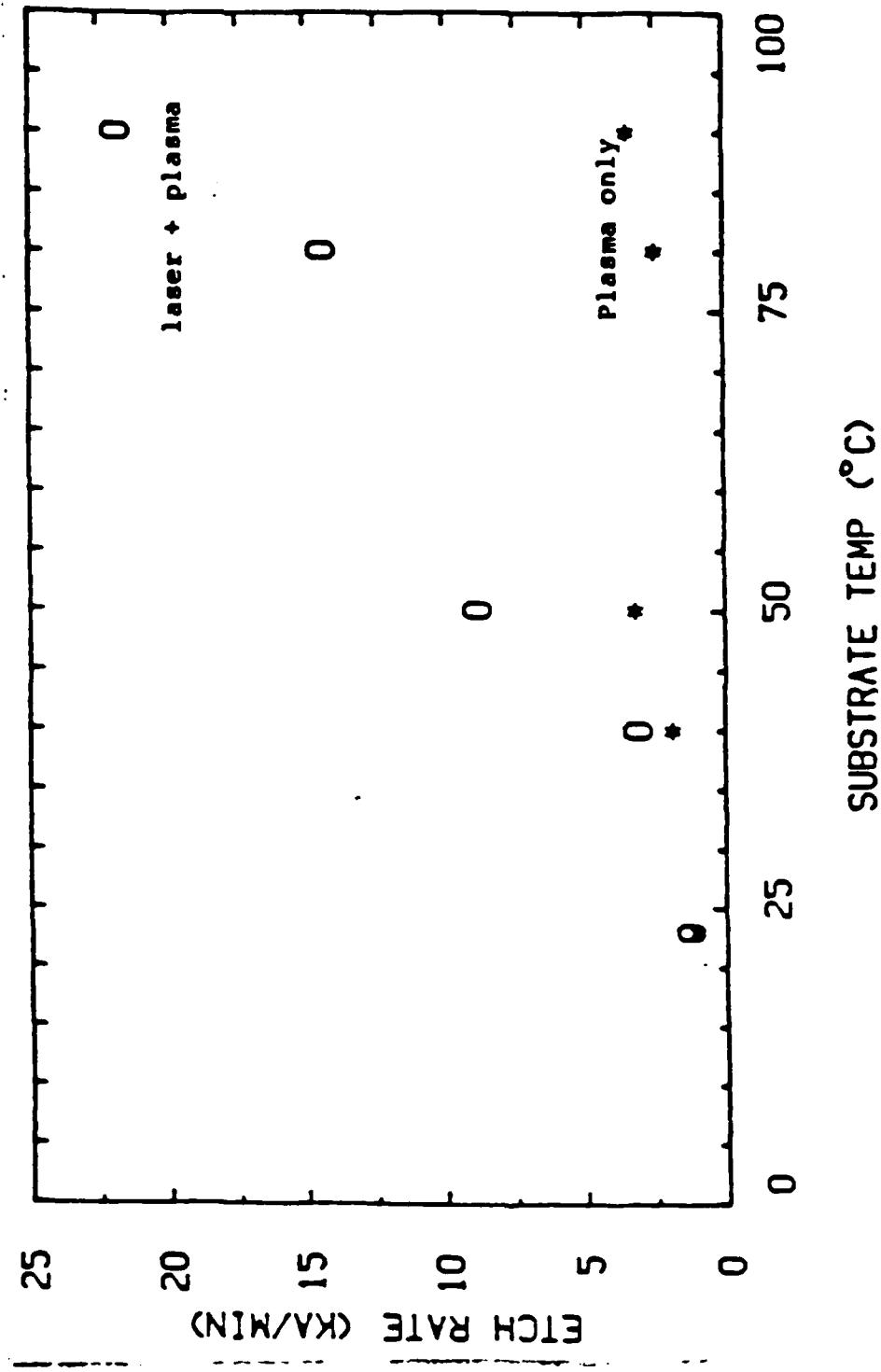
Figure 5: GaAs etch rate with plasma discharge alone and with plasma discharge and laser fluence together. The ambient gas is at 5 torr HBr, flowing at 25 sccm.

PLASMA CURRENT VS LASER POWER
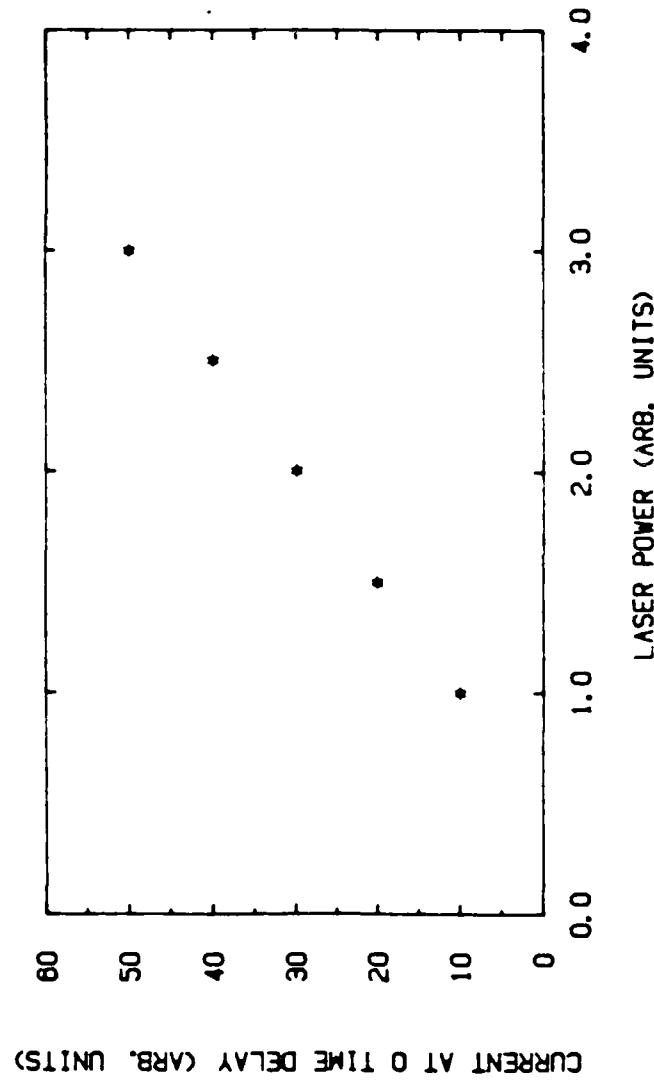
Figure 6: The influence of 193-nm light on the discharge current of a pulse d.c. discharge of HBr. The laser energy at 1.0 on the x-axis is approximately 2.5 mJ/cm². The laser pulse was not delayed with respect to the beginning of the discharge voltage.

OPTICALLY-INDUCED, ROOM-TEMPERATURE OXIDATION OF GALLIUM ARSENIDE

Chien-Fan Yu, Michael T. Schmidt, Dragan V. Podlesnik, and Richard M. Osgood Jr., Microelectronics Sciences Laboratories, Columbia University, New York, NY 10027

ABSTRACT

Room-temperature, optically-induced oxidation of the gallium arsenide surface has been studied with laser radiation of different wavelengths. It was found that deep-ultraviolet light is much more effective in enhancing oxidation than near-ultraviolet or visible light. The growth rate of the oxide was also found to be drastically increased by the presence of chemisorbed water molecules on the surface.

INTRODUCTION

The study of optically-induced oxidation of the gallium arsenide surface has been a subject of intensive investigation[1-5]. These studies have shown that simultaneous exposure of a semiconductor surface to oxygen and to laser light can stimulate the uptake of oxygen. It is also known that a thin oxide layer formed on the gallium arsenide surface can be used to passivate the surface or to alter the interface properties[6]. A better understanding and control of this room-temperature process can lend itself to many applications in the fabrication of semiconductor devices. The enhancement of oxidation on the gallium arsenide surface through above-band-gap photon illumination has been generally attributed to the effect of photogenerated carriers[1-5]. We report here our recent study on the wavelength dependence and the effect of water molecules adsorbed on the gallium arsenide surface in this photon-induced process. The use of deep ultraviolet light, below the oxygen dissociation limit, and/or the presence of chemisorbed water molecules on the surface were found to increase significantly the growth rate of the oxide.

EXPERIMENTAL

Experiments were carried out with either an argon-ion laser tuned to one of its visible or near-uv lines, or with a frequency-doubled argon-ion laser to generate 257-nm deep uv light. An excimer laser was also used as an alternative to generate 248-nm (KrF) light. In all cases the dissociation limit of oxygen molecules (242 nm) was not exceeded. The laser power was kept sufficiently low so that the temperature rise on the sample during exposure was negligible. The typical cw power density used was 50 mW/cm^2.

The gallium arsenide samples used in this experiment were n-type of (100) orientation with $\sim 10^{18}$ atoms/cm^3 doping density. Surfaces were cleaned by the following steps before ex-situ experiments: degreasing in warm trichloroethylene, immersion in acetone, immersion in methanol, immersion in DI water, and 50% NH_4OH rinse. Samples were then blown dry and attached to the mount with silver paste. For in-situ experiments, performed inside an UHV chamber, additional surface cleaning was done by heating to 540°C, followed by 500 eV argon-ion bombardment and then 540°C annealing for 5 minutes to reduce sputtering damage.

Surface analysis was performed using a multiprobe surface spectroscopy system equipped with a concentric hemispherical energy analyzer. All XPS measurements were carried out with an AlK$_\alpha$ (1486.6 eV) excitation source at 240 W. The system was calibrated from the Au $4f_{7/2}$ level with a binding

energy of 83.8 eV. The absence of surface charging was determined by observing the position of the C 1s level at 284.6 eV from the adventitious carbon on the sample. The pressure in the analysis chamber was kept below 5×10^{-10} Torr during the measurement.

RESULTS

To review, in earlier studies[7,8] we showed by using AES depth profiling that the oxide layer formed during laser irradiation on the gallium arsenide surface immersed in pure water is more than three times thicker than that made by irradiation in ambient air. However, while the oxide is arsenic deficient when made in water, it is nearly stoichiometric when made in air. In addition, the 257-nm laser light exposure was found to enhance the oxidation much more strongly than the 514-nm laser light of the same power density.

The XPS spectra showed that the oxide layer contains both gallium oxide and arsenic oxide. The As3d peak consists of well resolved components from the oxide and the substrate as shown in Fig. 1.
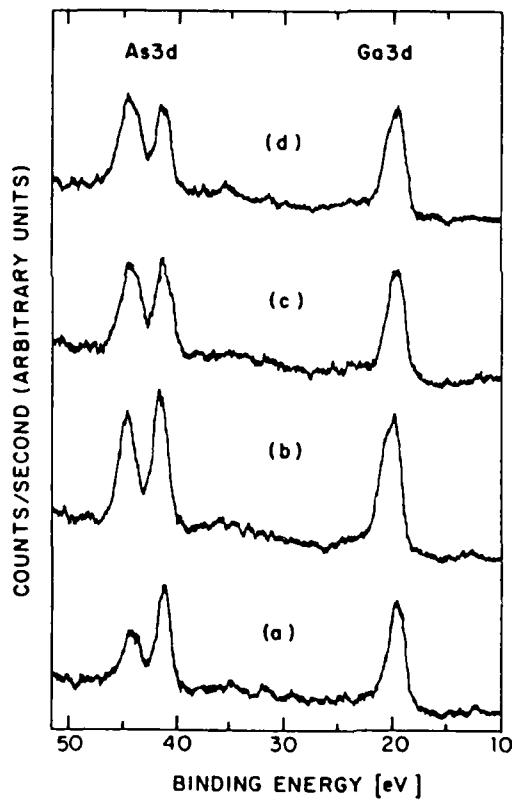

Figure 1. XPS spectra of the gallium arsenide surfaces irradiated with 248-nm laser light of 3 mJ/cm^2 at 30 Hz in laboratory air ambient for a) 10 min, b) 20 min, c) 32 min, and d) 40 min.

The thickness of the oxide layer can be estimated using the relative intensities of the oxide and substrate of the same emission level in the XPS spectrum as shown in the following equation[9]:

$$d_{ox} = \lambda_{ox} * \sin\theta * \ln(\frac{I_{ox}}{I_{sub}} * \frac{D_{sub}}{D_{ox}} * \frac{\lambda_{sub}}{\lambda_{ox}} + 1)$$

where I_{ox} and I_{sub} are the intensities of the photoelectrons from the same atomic level, D_{ox} and D_{sub} the densities of oxide and substrate, λ_{ox} and λ_{sub} the electron mean escape depths of oxide and substrate, d_{ox} the thickness of the oxide layer, and θ the exit angle of the photoelectrons relative to the surface. The electron mean escape depths in gallium arsenide have been determined[10] to be 25 Å for the As3d level and 7 Å for the As2p$_{3/2}$ level. We thus determined the thickness of the oxide layers shown in Fig. 1 to be 13, 17, 21, and 26 Å, made by exposure of GaAs surfaces to 248-nm laser light of 3 mJ/cm^2 at 30 Hz for 10, 20, 32, and 40 minutes, respectively. In the control sample, 6 Å oxide was found to have developed through the handling in air after the chemical cleaning procedures. Subtracting this native oxide thickness from each of the above thicknesses, a linear oxide growth rate of about 0.5 Å/min can be seen. A more extensive study of the wavelength dependence was also made. As shown in Fig. 2, exposure of the deep ultraviolet laser light, either 257-nm (cw) or 248-nm (pulsed), can make much more oxide than that of the near ultraviolet or visible laser light of about the same power intensity (ca. 50 mW/cm^2). Note the latter longer wavelengths induce only a small amount of additional oxide on top of the native oxide shown in the open circle.
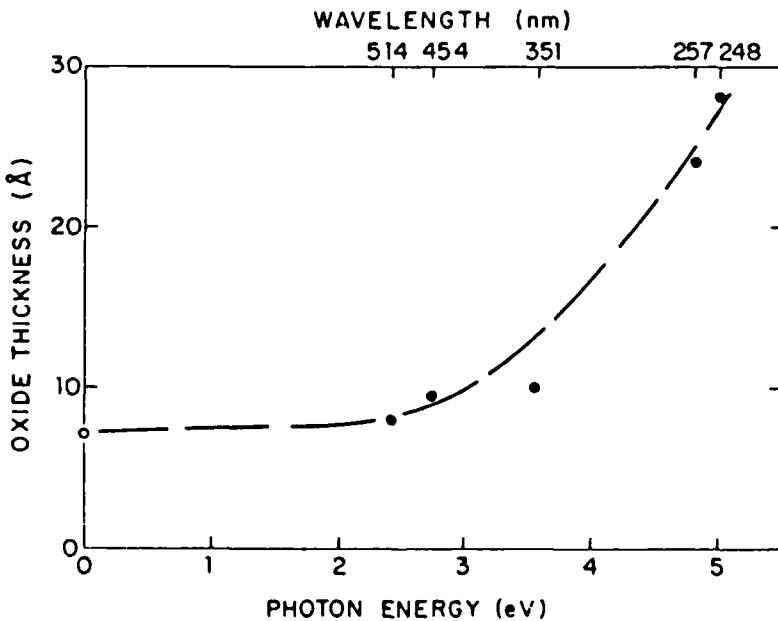

Figure 2. Oxide layer thickness as a function of the wavelength of the laser light irradiated at 50 mW/cm^2 for 20 minutes each in laboratory air ambient.
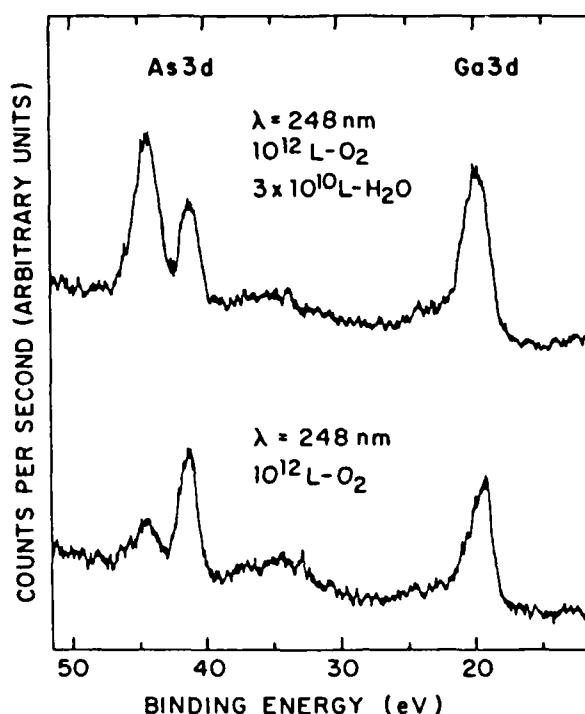
Figure 3. XPS spectra of gallium arsenide surfaces
irradiated with 248-nm laser light of 2 mJ/cm^2 at 100 Hz
for 20 minutes in dry oxygen (lower panel) and in oxygen
with equilibrium water vapor (upper panel).

Another set of experiments were done in the well-controlled gas ambient
inside an UHV chamber. The 248-nm light was introduced through a suprasil
window onto the sputter-cleaned and annealed surface inside the chamber. A
drastic difference was found depending on whether this chamber was filled
either with dry oxygen or with oxygen bubbled through pure water. Results
are shown in Fig. 3 for 20 minute exposure of 200 mW/cm^2 power density. In
the lower panel the oxide grown in dry oxygen is shown to be only about 9 Å
in thickness, while in the upper panel the oxide grown in water-vapor-
saturated oxygen is shown to be about 31 Å in thickness. In a separate
control experiment, oxygen-free water vapor (obtained by prolonged bubbling
with nitrogen) was let into the chamber up to its equilibrium vapor pressure
(∿ 20 torr). Subsequent irradiation of the surface with 248-nm light
produced no appreciable oxide.

DISCUSSION

We have not yet developed a complete understanding about the role the
carriers play in this optically-induced oxidation process. However,
unrelaxed carriers generated by the deep-uv light illumination are more
energetic than those generated by the near-uv or visible light. This may
explain, at least in part, the strong wavelength dependence of the oxidation
process.

Moreover, based on the previous understanding of GaAs surface oxidation, we can develop a plausible explanation for the results of the enhanced oxidation through the presence of water adsorbates. It has been reported that the adsorption of water molecules on the gallium arsenide surface at room temperature will lead to a chemisorbed adlayer on the surface[11,12]. This may then be followed by the dissociation of the water molecule into chemically bonded -OH and -H to surface gallium atoms[12]. The oxidation of gallium arsenide is believed to proceed with the initial attachment of oxygen to arsenic atoms[5,13]. We postulate that the chemisorption of water molecules on the surface will weaken the adjacent Ga-As bonds and make the arsenic atoms more susceptible for chemisorption of oxygen molecules. Photogenerated carriers will subsequently induce the breaking of the molecular oxygen bonds and complete the oxidation process. Thus, a much more rapid growth of oxide is observed with the presence of water vapor.

ACKNOWLEDGMENT

REFERENCES

1. S.A. Schafer and S.A. Lyon, J. Vac. Sci. Technol. 19, 494 (1981).
2. W.G. Petro, I. Hino, S. Eglash, I. Lindau, C.Y. Su and W.E. Spicer, J. Vac. Sci. Technol. 21, 405 (1982).
3. V.M. Bermudez, J. Appl. Phys. 54, 6795 (1983).
4. F. Bartels and W. Mönch, Surf. Sci. 143, 315 (1984).
5. K.A. Bertness, W.G. Petro, J.A. Silberman, D.J. Friedman, and W.E. Spicer, J. Vac. Sci. Technol. A 3, 1464 (1985).
6. S.D. Offsey, J.M. Woodall, A.C. Warren, P.D. Kirchner, T.I. Chappell and G.D. Pettit, Appl. Phys. Lett. 48, 475 (1986).
7. D.V. Podlesnik, H.H. Gilgen, A.E. Willner and R.M. Osgood Jr., J. Opt. Soc. Am. B 3, 775 (1986).
8. C.F. Yu, D.V. Podlesnik, M.T. Schmidt, H.H. Gilgen and R.M. Osgood Jr., Chem. Phys. Lett. 130, 301 (1986).
9. T.A. Carlson and G.E. McGuire, J. Electron Spectrosc. 1, 161 (1972/73).
10. H. Gant and W. Mönch, Surf. Sci. 105, 217 (1981).
11. M. Büchel and H. Lüth, Surf. Sci. 87, 285 (1979).
12. C. Webb and M. Lichtensteiger, J. Vac. Sci. Technol. 21, 659 (1982).
13. G. Langren, R. Ludeke, Y. Jugnet, J.F. Morar, and F.J. Himpsel, J. Vac. Sci. Technol. B 2, 351 (1984).

Large Area Laser-Assisted Etching of Electronic Materials

Peter D. Brewer

Hughes Research Laboratories
Malibu, California
90265

and

R. M. Osgood, Jr.

Columbia University
Microelectronics Sciences Laboratories
New York, New York
10027

Abstract

The use of laser assisted-chemistry for dry etching of electronic materials is described
Emphasis is placed on the use of laser-assisted reactions for large area processing   Review
of the current technology is given for large area masked etching, UV-projection etching, and
laser assisted reactive ion etching (RIE), and plasma etching

Introduction

Many new laser-chemical techniques for dry etching have been developed for potential
application in microelectronic fabrication   The applications range from high resolution,
planar, UV-projection pattern transfer to maskless fabrication of three dimensional micro-
structures   Most of these techniques have in common the use of laser-induced or -enhanced
chemical reactions at a solid-gas or solid-liquid interface. In this paper, we will review
the status of large area laser-assisted etching of materials for solid state electronics
with emphasis on gas (dry) ambients   The reader interested in "direct write" etching and
basic mechanisms involved in laser assisted etching is referred to several recent reviews.[1]

The features which make laser-assisted etching attractive from a processing perspective
are the fast etch rates, specific chemistry, low substrate temperatures, high material
selectivity, and etch anisotropy. These features are a result of the nature of the laser-
initiated chemistry involved   Specifically, it is possible to selectively drive chemical
reactions along predetermined pathways using laser excitation of specific reaction modes
For example, using laser-induced chemistry, it may be possible to produce CF_3 and Br reac-
tants from CF_3Br without the concomitant generation of other species (CF, C, CF_3^+, etc),
such as those produced in plasma discharges [2] In other forms of laser processing, chemical
reactions are driven by rapid transient heating of the substrate by a pulsed laser   By
using a short-pulsed laser with a small surface penetration depth, the heat is confined only
to the surface   Consequently, high temperatures can be reached in the near-surface region
where the remainder of the sample remains at the ambient temperature   In addition, the
optical characteristics of the laser beam itself can give the etching process desirable
properties   For example, since the laser beam has a well defined direction, highly
anisotropic features can be produced   Finally, focusing or imaging the laser beam enables
unique maskless pattern generation on surfaces by the techniques of "direct writing" or
projection patterning, respectively

We will first briefly review the general field of laser-assisted etching of electronic
materials and then review the current status of large area masked laser etching, laser-
assisted projection etching and laser-enhanced RIE and plasma etching

An Overview of Chemical Approaches to Laser Etching

Chemical approaches to laser-assisted etching of electronic materials can be treated in
terms of four general mechanisms according to the involvement of the laser light in several
basic solid-surface and gas phase molecular processes   These mechanisms are photochemical,
thermal, photon-activated surface, and ablative. For example, in photon-assisted dry
etching the process is dominated by photon interactions with gas-phase or surface-adsorbed
species, while in direct ablative photoetching photon-surface processes dominate the
material removal   Table I categorizes dry laser etching processes in terms of these four
mechanisms

Table I

Current Summary of Laser-Assisted Dry Etching of Electronic Materials

Mechanism	Material	Etchant Gas	Laser Source	Comments	Ref
Photochemical	Si	Br_2	$Ar^*/\|\|$	Gas Activated	20
	Si, Ta	SF_6	Pulsed $CO_2/\|\|$	Vibrationally Activated	1A
	Si	Cl_2/HCl	$Ar^*/\perp$	Gas Activated Some Surface Heating	19
	Si/poly Si	$Cl_2+Si(Ch_3)_4$ Cl^*+MMA	$XeCl/\perp$	Gas Activated Surface Protected Mechanism	10, 11
	Ge	Br_2	$Ar^*/\|\|$	Gas Activated	21
	GaAs	CF_3Br, CH_3Br, HBr	$ArF/\perp$	Gas Activated	8,9
	GaAs/InP	CH_3Br, CH_3Cl, CF_3I	Doubled $Ar^*/\perp$	Gas Activated	22
	SiO_2	CF_3Br, CDF_3	Pulsed $CO_2/\|\|$	Multiphoton Gas Phase Dissociation	23
	SiO_2	Cl_2	$Ar^*/\perp$	Gas Activated	24
	SiO_2	CCl_2F_2	$KrF/\perp$		25
	SiO_2	NF_3+H_2	$ArF/\perp$		26
	SiO_2/B_2O_3	CF_2Br_2	$KrF/ArF/\perp$	Gas Activated	27
	Si_3N_4	F_2CO, NF_3	$ArF/\perp$	Gas Activated	25
	W, Mo, Ti	F_2CO, NF_3	$ArF/\perp$	Gas Activated	14
Surface Activated (Nonthermal)	Si	XeF_2	$Ar^*/\perp$	Reaction Enhanced by Electron-Hole Pair Formation	6
	GaAs/GaP	Cl	$Ar^*/\perp$	Band Bending Under Illumination and Biasing	28
	Si	XeF_2	Pulsed CO_2	Some Thermal Effects, Photons Increase XeF_2 Sticking Probability	29
Thermal	Ge	Br_2	Vis Dye Laser/$\perp$		30
	GaAs	Cl_2	$Ar^*/\perp$	Formation of Via Holes	31
	Ge, Si, GaAs, ZnSe	CF_3I, Br_2	cw $CO_2/\perp$		7
	GaAs, InP, InSb	CCl_4, $SiCl_4$, $GeCl_4$	$Ar^*/\perp$		32
	SiO_2	HF	cw CO_2		33
	SiO_2	HF, HCl, H_2, NF_3, CF_3Br, Cl_2	CO_2		34
	Pyrex	H_2	ArF		12
	Ceramics	Cl_2	Ar^*		35
	Mo, W, Si	F_2/He	cw $CO_2/\perp$		36
	MnZn Ferrites	CCl_4, CF_3Cl	Ar^*		1a
	$Ni_{0.2}Fe_{0.8}$	SF_6, CF_4, CCl_4, Cl_2	Ar^*		1a
Direct Ablative	PMMA	Air	Excimers		5
	Polyimide	Air	Excimers		36
	Nitrocellulose	Air	Excimers		38
	Al	Cl_2	Excimers		4
	Au	Air	XeCl		3

In the case of photochemical dry etching, an inert parent molecule in the vicinity of a gas-solid interface is dissociated by one of several gas or adsorbed phase processes to produce a reactive atom or free radical. This species reacts with the surface to form a volatile compound which then leaves the surface, thereby resulting in etching of the solid. In the simplest case, absorption of a single UV photon results in dissociation of the parent molecule to produce reactants. Alternatively, multiphoton excitation by a pulsed CO_2 laser has been used to highly excite and/or dissociate the gas phase parent molecules.[1a]

For direct ablative photoetching, the laser wavelength is selected such that the light is highly absorbed in the surface layer of the substrate. Using lasers which deliver pulses with high peak powers, the surface layer is removed by simple evaporation. This approach has been applied to a wide variety of inorganic materials.[3] In a novel variation of this approach, Koren et al.[4] have etched thin Al films which are prereacted with molecular chlorine to produce $AlCl_3$. The surface layer was then removed by XeCl (308 nm) laser-induced evaporation. This procedure has been used to etch Al films at a rate of 1 μm/pulse. On the other hand, for etching of polymers with UV lasers, explosive removal of surface layers is believed to be caused by bond cleavage photochemistry in the surface molecules due to direct polymer electronic transitions.[5]

For etching semiconductor substrates irradiated with low laser powers in the presence of a reactive ambient, electron-hole pairs created in the bulk by absorption of above band gap laser light in the bulk have been shown to be responsible for accelerated desorption and reactions at the semiconductor surface.[6] At higher intensities, thermal excitation of the solid becomes important. The heating results from a number of phenomena, including non-radiative recombination and energy transfer from hot electrons. The thermal excitation is needed to drive chemical reactions over an energy threshold on the surface.[7] Parent molecules are then selected which will only react with the local heated surface.

UV-Initiated Dry Etching of Masked Substrates

Many semiconductors materials such as GaAs are particularly difficult to etch in plasma environments because of the surface damage created by energetic charged particles. This is also true of etching of substrates having thin oxide layers or structures. In order to overcome these problems while keeping the desirable anisotropic characteristics of plasma etching, UV-initiated dry etching is being developed as a low energy, neutral charge technique having applications to large area processing. Excimer laser dry etching has been demonstrated for the masked etching of GaAs, poly-Si and Si.[8-11] In the case of GaAs,[8,9] an ArF excimer laser has been employed to effect single photon dissociation of either CF_3Br, CH_3Br, or HBr parent gases to produce primarily Br atom etchants. For both poly-Si and Si, mixtures of the etchant parent gas and a polymerizing agent were employed to etch the substrate while producing anisotropic side-wall features.

Figure 1 shows the etch rate of GaAs (100) as a function of ArF laser energy density at two temperatures when HBr is used as the parent gas. Masking of the GaAs substrates was accomplished using photoresists which do not ablate during the excimer laser pulse. Line-width resolution below 1 μm has been demonstrated.[8] Etch rates of up to 8 μm/min with large area substrates heated to 60°C were obtained using moderate laser intensities (<35 mJ/cm²).[9] Selective crystallographic etching was observed (Figure 2) and etch rates for the lower order crystal planes were measured to be {111}B>{100}>{110}>{111}A. The etching process seems to be dominated by reactions between Br radicals and the GaAs surface, etching products are primarily the bromides of gallium and arsenic. In these experiments, no evidence of As depletion of the surface layer has been observed.

For the masked excimer laser initiated etching of poly-Si[10] and Si,[11] Horiike and co-workers at the Toshiba Corp. have developed a novel self-masking technique to control side-wall profiles during processing by using etchant gas/monomer mixtures irradiated with an excimer laser. In their initial work,[10] a XeCl excimer laser (308 nm) is used to photo-dissociate a mixture of Cl_2 and $Si(CH_3)_4$. The Cl atoms generated by the photodissociation process served both to etch the substrate and to cause free radical polymerization of the $Si(CH_3)_4$ monomer. Polymerization produces a nonvolatile poly-methylated Si surface film. Direct illumination by the XeCl laser normal to the substrate causes ablation of the deposited film. This results in anisotropic etching since the sidewalls (parallel to the laser beam) remain protected during etching. The chlorine chemistry produces selective etching of the Si or poly-Si relative to the SiO_2 masking material. As a result, the process is potentially useful for the etching submicrometer features with very thin gate oxides. Note, however, that care must be taken with UV illumination to prevent photogenerated charge from being produced under the gate oxide. An alternative approach to this technique has been reported.[11] In these experiments, a microwave discharge was used to produce the etchant Cl radicals, and methyl-methacrylate (MMA) was used as the polymer precursor. These changes enabled the decoupling of the etchant production from the ablation of the polymerized surface films. The implementation of this technology in the fabrication of a 1 Mbit DRAM cell has been reported.[11]

GaAs (100) ETCH RATE VS LASER POWER
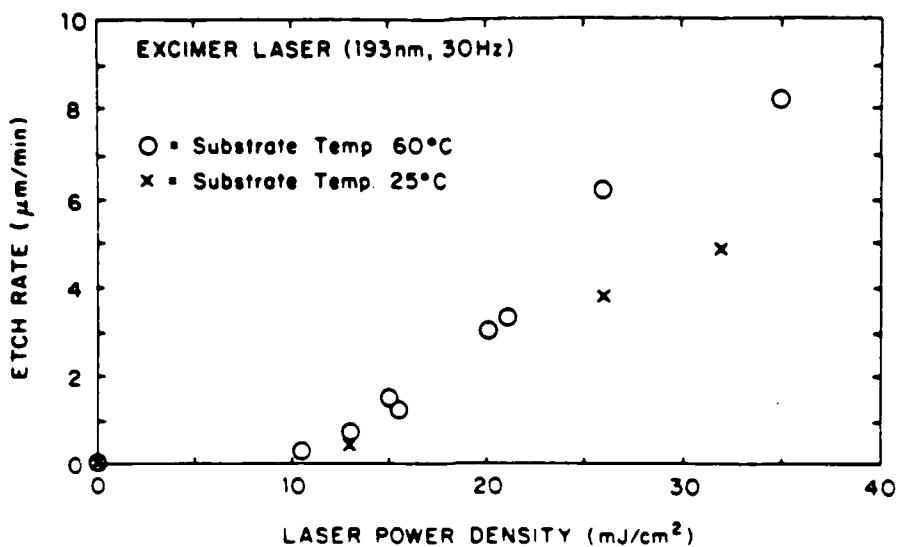
Figure 1   The etch rate of GaAs (100) as a function of laser energy density is plotted for two sample temperatures. The laser beam was oriented at normal incident to the surface.
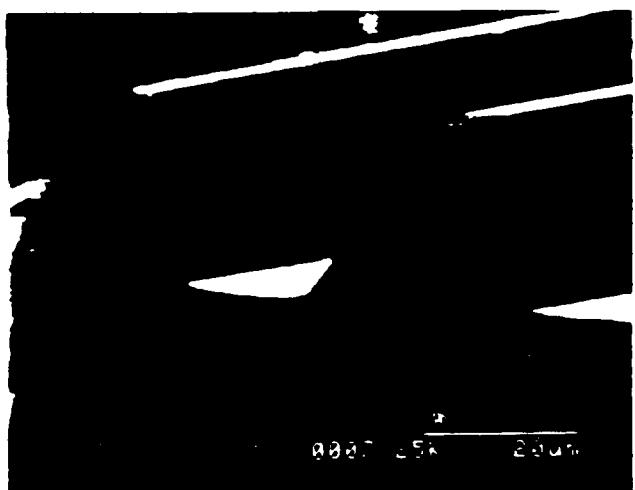
Figure 2   Scanning electron micrograph of an GaAs (100) sample etched using HBr gas and 193 nm excimer laser radiation (30 Hz, 15 mJ/cm²) The crystal was masked with photoresist lines oriented normal to the (110) cleavage plane Crystal planes produced during etching are indicated in the micrograph

UV-Projection Etching

Projection etching has significant advantages for the fabrication of microelectronic circuits since the characteristics of laser processing are combined in both a parallel and a direct pattern transfer technique. Pattern transfer has been accomplished by imaging a homogenized laser beam through a photomask onto a solid surface to produce patterned micro-chemical reactions. This scheme has been applied to the etching of pyrex glass,[12] GaAs,[13] and poly-Si,[11,14] although recently surface product ablative patterning of Al films[4] has been demonstrated in a manner similar to that used for etching polymers.[6] In work on Mo and poly-Si,[14] radiation present in near-field diffraction patterns has been used to photochem-ically initiate etching reactions without substrate masking. A resolution of <0.30 μm was achieved using this process. In all of these cases, excimer lasers have been used as the illumination source. The primary advantages of these sources are their short wavelengths and high average and peak powers. Relatively simple optical imaging systems have been employed with typical experimental setups using standard photomasks and 1:1 imaging lenses. For higher resolution reflective microscope systems (35.1 reduction) have been used[12,13] to achieve etch features of <1 μm.

For the projection etching of GaAs,[13] poly-Si,[11,14] and Mo,[14] a gas-phase photochem-ically initiated process relying on the dissociation of halogen-bearing parent gases was employed. The physical considerations in these experiments are different from those of masked substrate etching since parent gas pressure, surface temperature, and laser intensity all affect the ultimate image resolution through gas or surface-mediated diffusion. In the case of GaAs, the addition of reactive and inert buffer gases[13] have been found to be useful in controlling the etch linewidth (Figure 3); however, this results in lower etch rates. In these experiments, the confinement of the etching gas diffusion appears to stem from the reduction of the mean free path by physical collisions or chemical reactions. However, compared to theoretical fits to the experimental parameters, the etched linewidths have been found to be smaller by a factor of 5. The order of the reaction of the Br atoms with the GaAs surface has been used to explain this large decrease in linewidth. Physically, this can be understood in terms of a concentration lifetime due to the nonlinear order of the surface reaction which limits the etching contribution of the initial projected reagent flux to less than one mean free path above the surface.[13] In other words, the fact that the etching is nonlinear in reactant flux means that the surface regions which encounter a high atom flux are etched much faster than the low concentration regions. The result of this effect is that the etched image more closely reproduces the initial projected image.

15935-1
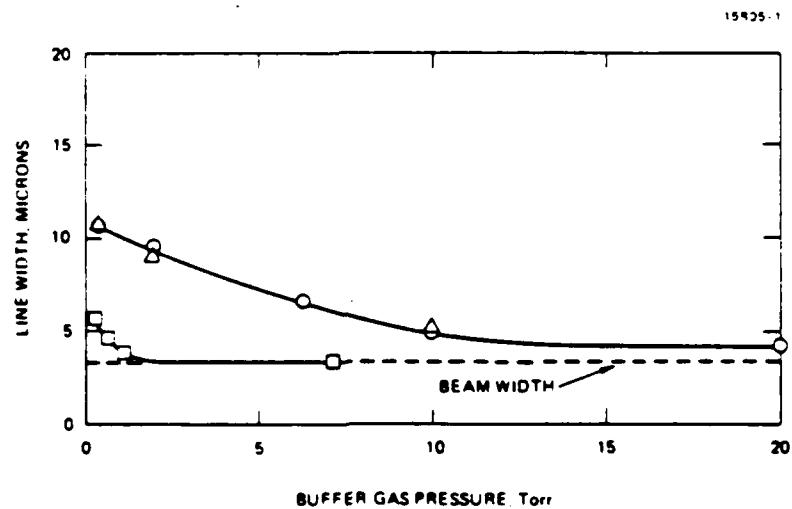
Figure 3   Linewidth of etched grooves are plotted as a function of buffer gas pressure. Two physical quenchers (Ar=0 and H$_2$=Δ) and a chemical scavenger (ethylene $\Box$) were mixed with the HBr parent gas. The optical feature size of the projected line image was 3.5 μm.

In other projection work, waveguided aqueous etching of GaAs has been extended to be a parallel process by the use of a UV projection scheme.[15] In these experiments, the light of a KrF excimer laser (248 nm) is projected onto a GaAs substrate which is placed in a quartz cell containing a 1 to 5% HNO_3 solution. An array of high aspect ratio holes (4 μm diameter) have been produced without masking the substrate as shown in Figure 4.
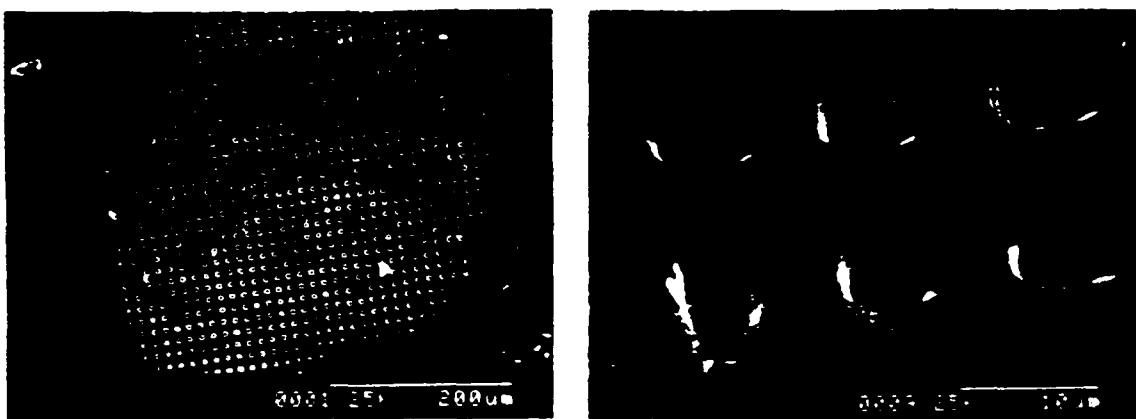

Figure 4. Scanning electron micrograph of laser-induced wet etched GaAs (100) using 5% HNO_3 and a KrF excimer laser. Projection imaging by 4:1 optical reduction was employed to produce these high resolution holes (4 mm diameter).

The above approaches are in contrast to the direct physical ablative etching of thin polymer[5] and Al[4] films where the laser is not used to initiate external chemistry. It is thought that in the case of polymer films at low laser powers that electronic transitions in the surface layer control the etching. For organic resists,[5] experiments have demonstrated <1 μm resolution. In addition, ablation has been achieved at laser energy densities as low as 10 mJ/cm^2. For ablative etching of Al films, the surfaces are exposed to molecular chlorine to form an $AlCl_3$ outer layer which is ablatively desorbed using a 0.25 to 3 J/cm^2 XeCl excimer pulse. Etch rates up to 1 μm/pulse have been achieved in these experiments. This approach is valuable for some materials; however, for electronically sensitive materials, such as compound semiconductors, direct physical ablation would be deleterious to the surface stoichiometry due to surface heating.

Laser Enhanced RIE and Plasma Etching

The use of lasers in conjunction with RIE and plasma etching reactors has only recently been reported.[16,17,18] In one approach, a port in the top electrode of a conventional reactor is used to introduce the laser light, which can be focused[16,17] or projected[18] by an external lens system onto the semiconductor surface. Localized patterns can be etched in the area illuminated by the laser. With the use of projection etching, moderate image resolution <10 μm has been achieved.[18] The limitation in feature size which has been achieved in this work is due to the simple lens system employed. A second approach uses the laser light to alter the external gas-phase chemistry above the semiconductor.[18] This allows the separate control of the production of gas-phase neutral chemical species and ion bombardment in the plasma etcher. In these experiments, the output of a 193 nm ArF excimer laser was directed over a GaAs sample in a pulsed discharge of Ar and HBr.

There are several potential advantages to using laser-controlled chemistry in conjunction with a plasma reactor. First, the light beam can be used to define a region to be etched without the necessity of prior surface masking. Second, light can enhance the etch rate of materials with otherwise slow etching rates in conventional reactors. In addition, light-induced desorption may reduce the number of surface-adsorbed product molecules without requiring intense physical sputtering.

The mechanisms in the direct surface excitation orientation have been shown to fall into two classes, namely thermal and nonthermal. Nonthermal effects observed at low laser intensities have been shown, in the case of silicon, to depend on dopant concentrations and laser wavelength.[16] The doping dependence suggests that the photo-enhancement is related to the production of photogenerated carriers. It has been suggested that a flow of minority carriers to the surface reaction sites gives rise to enhanced surface reaction rates. This has

also been observed in laser enhanced etching of Si in XeF_2 (no plasma)[6] and in electrode-less laser-induced wet etching of GaAs.[16] At higher laser intensities, the local temper-ature may reach several hundred degrees at the center of the laser focal spot on the wafer depending on the focal spot size and the laser power. Etch rates of 2.5 times above the dark etch rate for Si[16] and 10 times for GaAs[17] have been obtained by localized heating of the semiconductor surface. This enhancement may result from product desorption, enhanced rates, or removing adlayers which block the surface reaction sites.

In the case of gas phase excitation by a 193 nm excimer laser in pulsed HBr/Ar dis-charges,[18] the densities of neutral species was controlled independently from the ion bom-bardment. Optical emission studies showed the increased production of reactive neutral radical species. An added advantage observed using this combination was that the plasma breakdown voltage was lower when the laser pulse was synchronized to the discharge pulse. This dependence also yielded lifetime information about the laser-produced radicals. Sub-stantial changes in the etching characteristics due to the laser chemistry were also observed with microstructure wall definition changing from vertical to crystallographic.

Conclusions

Laser-assisted etching has been shown to be a powerful method for processing electronic materials in regimes where plasma or RIE techniques fail. The most attractive features of laser-assisted etching include: maskless processing, etching with neutral species, and pro-duction of anisotropic features. For most materials, laser etching has been shown to have superior etch rates and material selectivity relative to plasma etching. Resolution of below 1 μm in both masked and projection laser etching has been demonstrated and has been found to be competitive with most other current processing techniques. The most important limitation to date is the need for more reliable, high power UV lasers.

Acknowledgments

References

1. a) Review of gas-surface chemistry in etching: Chuang, T. J., J. Vac. Sci. Technol., Vol. 21, p. 789 1982. b) Review of "direct write" etching. Ehrlich, D. J. and Tsao, J. Y., VLSI Electronics Microstructure Science, Vol. 7, p. 129 1983. c) General review of lasers in microelectronics. Osgood, Jr., R. M., Ann. Rev. Phys. Chem. Vol. 34, p. 77 1983.

2. Flamm, D., J. Appl. Phys., Vol. 51, p. 5688 1981.

3. Andrew, J., Dyre, P., Greenough, R., and Key, P., In Proceedings of the Conference on Lasers and Electro-Optics, Baltimore, MD 1983.

4. Koren, G., Ho, F., and Ritsko, J. J., Appl. Phys. Lett., Vol. 46, p. 1006 1985.

5. Srinivasan, R., and Mayne-Banton, V., Appl. Phys. Lett., Vol. 41, p. 576 1982. Latta, M., Morre, R., Rice, S., and Jain, K., J. Appl. Phys., Vol. 56, p. 586 1984.

6. Houle, F. A., S.P.I.E. Laser Assisted Deposition, Etching and Doping, Vol. 459, p. 110 (1984). Houle, F. A., Chem. Phys. Lett., Vol. 95, p. 5 1983. Houle, F. A., J. Chem. Phys., Vol. 79, p. 4237 1983.

7. Karlov, N. V., Luk'yanchuk, B. S., Sisakyan, E. V., and Shafeev, G. A., Sov. J. Quantum Electron., Vol. 15, p. 522 1985.

8. Brewer, P. D., Halle, S., and Osgood, Jr., R. M., Appl. Phys. Lett., Vol. 45, p. 475 1984.

9. Brewer, P. D., McClure, D., and Osgood, Jr., R. M., Appl. Phys. Lett., Vol. 47, p. 310 1985.

10. Horiike, Y., Sekine, M., Horioka, K., Arikado, T., Nakase, M., and H. Okano, H., Mat. Res. Soc. Ext. Abs., Eds. Houle, F. A., Deutsch, T. F., Osgood, Jr., R. M., p. 99 1984.

11. Okano, H., Hayasaka, N., Suto, S., Sekine, M., and Horiike, Y., Mat. Res. Soc. Ext. Abs., Eds. von Gutfeld, R. J., Greene, J. E., and Schlossberg, H., p. 121 1985.

12. Ehrlich, D.J., Tsao, J.Y., and Bozler, C.O., *J. Vac. Sci. Technol.*, Vol. B3, p. 1. 1985.

13. Brewer, P.D., McClure, D., and Osgood, Jr., R.M., *Appl. Phys. Lett.*, In Press.

14. Loper, G.L., and Tabat, M.D., *J. Appl. Phys.*, Vol. 58, p. 3649. 1985. Loper, G.L., and Tabat, M.D., *Appl. Phys. Lett.*, Vol. 46, p. 654. 1985.

15. Podlesnik, D.V., Gilgen, H.H., Brewer, P.D., and Osgood, Jr., R.M., In Press. Podlesnik, D.V., Gilgen, H.H., Osgood, Jr., R.M., *Appl. Phys. Lett.*, Vol. 43, p. 1083. 1983.

16. Holber, W., Reksten, G., and Osgood, Jr., R.M., *Appl. Phys. Lett.*, Vol. 46, p. 201. 1985.

17. Tsukada, N., Semura, S., Saito, H., Sugata, S., Asakawa, K., Mita, Y., *J. Appl. Phys.*, Vol. 55, p. 3417. 1984.

18. Holber, W., Chu, J., Flynn, G., Brewer, P.D., and Osgood, Jr., R.M., American Vacuum Society meeting Houston, Texas, Nov. 1985.

19. Ehrlich, D.J., Osgood, Jr., R.M., and Deutsch, T.F., *Appl. Phys. Lett.*, Vol. 38, p. 1018. 1981.

20. Beterov, I.M., Chebotaev, V.P., Yurshina, N.I. and Yurshin, B.Y., *Sov. J. Quantum Electron.*, Vol. 8, p. 1310. 1978.

21. Baklanov, M.R., Beterov, I.M., Repinskii, S.M., Rzhanov, A.V., Chebotaev, V.P., Yurshina, N.I., *Sov. Phys. Dokl.*, Vol. 19, p. 312. 1974.

22. Ehrlich, D.J., Osgood, Jr., R.M., and Deutsch, T.F., *Appl. Phys. Lett.*, Vol. 36, p. 698. 1980.

23. Steinfeld, J.I., Anderson, T.G., Reiser, C., Denison, D.R., Hartsough, L.D. and Hollahan, J.R., *J. Electrochem. Soc.*, Vol. 127, p. 514. 1980.

24. Houle, F.A., and Chuang, T.J., *J. Vac. Sci. Technol.*, Vol. 20, p. 790. 1982.

25. Loper, G.L., and Tabat, M.D., see Ref. 11, p. 133. 1985.

26. Yokoyama, S., Yamakage, Y., and Hirose, M., *Conf. on Solid State Devices and Materials Ext. Abs.*, Kobe, p. 451. 1984.

27. Brannon, J., see Ref. 10, p. 112.

28. Ashby, C.H.I., *Appl. Phys. Lett.*, Vol. 45, p. 892. 1984. Ashby, C.H.I., *Appl. Phys. Lett.*, Vol. 46, p. 752. 1985.

29. Chuang, T.J., *J. Chem. Phys.*, Vol. 74, p. 61. 1981.

30. Davis, G., Moore, C., and Gottscho, R., *SPIE*, Vol. 459, p. 115. 1984.

31. Tucker, A.W., and Birnbaum, M., *IEEE Device Lett.*, EDL-2, 39 (1983), McNevin, S.C., see Ref. 11, p. 135.

32. Takai, M., Tsuchimoto, J., Nakai, H., Tokuda, J. Gamo, K., and Namba, S., see Ref. 11, p. 129. 1985.

33. Daree, K., and Kaiser, W., *Glass Technol.*, Vol. 18, p. 19. 1977.

34. Dai, B.T., Agrawalla, B.S., and Allen, S.D., see Ref. 11, p. 143. 1985.

35. Johnson, A.W., and Smilgys, R.V., see Ref. 10, p. 108. 1984.

36. Koren, G., *Appl. Phys. Lett.*, Vol. 47, p. 1012. 1985.

37. Brannon, J.H., Lankard, J.R., Baise, A.I., Burns, F. and Kaufman, J., *J. Appl. Phys.*, Vol. 58, p. 2036. 1985.

38. Geis, M.W., Randall, J.N., Deutsch, T.F., Efremow, N.N., Donnelly, J.P., and Woodhouse, J.D., *J. Vac. Sci. Techno.*, Vol. B1, p. 1204. 1983.

DIRECT CONNECTION OF OPTICAL FIBERS TO INTEGRATED CIRCUITS

R.W. Ade, A. Amirfazli, T. Cacouris
E.R. Fossum, P.R. Prucnal, and R.M. Osgood

Department of Electrical Engineering and
NSF Center for Telecommunications Research
1312 S.W. Mudd Building
Columbia University
New York, New York 10027
(212) 280-3107

ABSTRACT

We have been investigating novel structures for coupling optical fibers directly to integrated circuits. The coupler features a small footprint, reliable alignment, and mechanical stability, and is suitable for high density optical fiber interconnects. At the present time, we have demonstrated a high quality (silicon) receiver structure fabricated using these techniques. The receiver consists of a vertical anisotropic cavity, formed using a laser-assisted etching process, which is then doped to form a P-N junction detector. A tapered optical fiber is inserted into the cavity and affixed in place by means of a U.V. cured adhesive. The performance of this receiver is significantly improved over that which we reported earlier this year[1]. The detector responsivity is 0.17 A/W (632.8 nm) and the dark current is less than 0.5 nA. We are presently making switching speed measurements, but the performance is expected to be comparable to other silicon P-N junction detectors.

The paper will briefly discuss the fabrication process of both the receiver and the optical fiber. The performance properties of the receiver as obtained at the time of the conference will be reported. We will also report our progress on backside interconnect structures, III-V device interconnects, and optical source structures, which are currently under investigation. Finally, the paper will address the problems we have encountered in mounting and packaging this device, and the implications of this sort of optical interconnection for the future.

1. P.R. Prucnal, E.R. Fossum, and R.M. Osgood, "Integrated Fiber-Optic Coupler for Very Large Scale Integration Interconnects", Opt. Lett., 11, 109, 1986.

Fig. 2. Photograph of etched single-mode fiber at 150x magnification.

Fig. 4. Dark and illuminated I-V characteristic of finished device (5μA/1V per div.).


Fig. 1. Schematic cross-section of interconnect device structure.

Fig. 3. Insertion of fiber into fabricated detector chip.

Wavelength dependence of optically induced oxidation of GaAs(100)

C. F. Yu, M. T. Schmidt, D. V. Podlesnik, and R. M. Osgood, Jr.

Microelectronics Sciences Laboratories, Columbia University, New York, New York, 10027-6699

The wavelength dependence of optically induced, room-temperature oxidation of GaAs (100) is reported. The sample illumination was performed with either continuous wave (cw) or pulsed-laser sources radiating at different wavelengths, in the range of 248–514 nm. The thickness of optically induced oxides, ranging from 1 to > 50 Å, was measured with x-ray photoelectron spectroscopy. It was found that the oxidation is always much more rapid in deep-ultraviolet than in near-ultraviolet or visible wavelength regions. This effect is attributed to the generation of hot carriers at the semiconductor surface under ultraviolet light.

I. INTRODUCTION

The enhancement of chemical reactions on semiconductor surfaces by illumination with photons is well known. In many cases, this enhancement occurs as a result of photogeneration of electron-hole pairs, with a subsequent step by one of the carriers leading to or catalyzing a chemical reaction at the semiconductor surface.[1-3]

Such reactions are of practical interest because of the potential for utility in the fabrication of semiconductor devices. For example, III–V and II–VI semiconductors often require processes that accelerate chemical reactions while keeping the semiconductor near room temperature. It has not been generally appreciated, however, that carrier-related reactions exhibit strong-band-structure dependence when different photon energies are used. In GaAs crystals, for example, visible and ultraviolet photons are absorbed in different parts of the Brillouin zone, thus generating electrons and holes of different energies. In this paper, we will present the results of a study of the light-enhanced oxidation of GaAs. It will be shown that the optical properties of GaAs at deep ultraviolet wavelengths lead to more rapid oxidation reactions than those seen in visible wavelengths. The enhanced ultraviolet oxidation was then used to grow relatively thick oxides (> 50 Å) at room temperature.

II. OPTICAL PROPERTIES OF GaAs

In our investigation of the laser-enhanced oxidation of GaAs, we are concerned with the optical properties of GaAs in the visible and ultraviolet wavelength range. Measurements of normal incidence reflectance,[4] and more recently, measurements using spectroscopic ellipsometry[5] have provided optical data for the intrinsic GaAs surface. The results for the wide spectral region 0–15 eV can be best summarized by the complex dielectric function. In the visible and ultraviolet region, 2–5 eV, the dielectric function[4,5] varies rapidly and is sharply structured. Philipp and Ehrenreich[4] commented on the spectral region at higher energies than the intrinsic absorption edge (the band gap). They noted that the sharp structures for hv between 2 and 5 eV were associated with the vertical, valence-to-conduction-band transitions in various parts of the Brillouin zone. For example, 3 eV photons induce a transition at the L-edge of the Brillouin zone.[6] i.e., in the [111] direction (see Fig. 1, after Ref. 7). Similarly, 5 eV photons induce an X-edge transition[6] in the

[100] direction. As a compliment to the dielectric function, the optical coefficient curve[8] can be extracted from the same measurements. The distinct absorption peaks of the absorption coefficient for photon energies near 3 and 5 eV can be associated with the above-mentioned valence to conduction-band vertical transitions in the Brillouin zone. An important effect is that going from the green to the deep ultraviolet, the absorption coefficient increases by a factor of more than 20. As a result, in the ultraviolet region, the photogeneration is confined to within a few monolayers of the GaAs surface, e.g., 50 Å for 250 nm photon illumination.

Photon absorption in different parts of the Brillouin zone results in the generation of electrons and holes of different energies. In GaAs crystals, visible light photons, $hv \sim 2.5$ eV, are absorbed by the electronic transitions in the center of the Brillouin zone. Inspection of the energy band diagram, Fig. 1, shows that the excess energy is transferred mainly to the electrons. However, ultraviolet photons, $hv \sim 5$ eV, induce a transition at the X-edge of the Brillouin zone, and holes now receive most of the excess photoenergy. Specifically, the ultraviolet-generated holes are created with energies of ~ 2.5
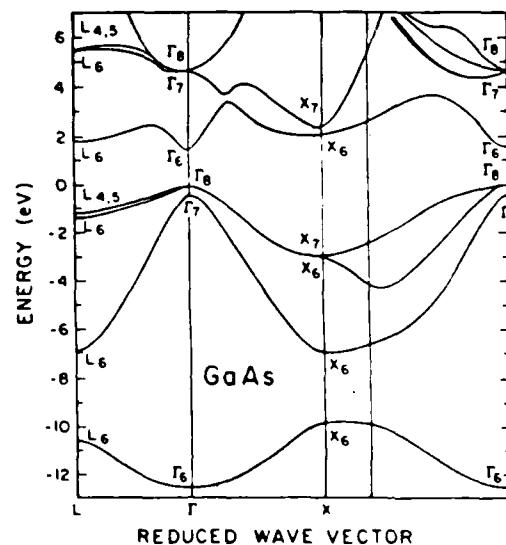

FIG. 1 Electron energy vs reduced wave vector for the four GaAs valence bands and the first several conduction bands, after Ref. 7

eV below the absolute top of the valence band. Notice that since these holes are produced spatially within 5 nm of the surface, energy relaxation is incomplete for carriers reaching the surface.[3]

A consequence of using these nonthermal carriers is that the interface chemistry is altered under ultraviolet illumination. In this work, we show that ultraviolet light is sufficiently actinic that rapid oxidation reactions occur, in contrast to visible light, for which under the same conditions, no effect of a comparable magnitude could be observed.

III. EXPERIMENT

Initial oxidation experiments were carried out with a continuous-wave (cw) argon-ion laser either tuned to one of its visible or near-ultraviolet lines, or with a frequency-doubled argon-ion laser to generate 257-nm deep ultraviolet light. In all cases the laser intensity was ~ 50 mW/cm^2. For the experiments carried out in the ultrahigh vacuum (UHV) system, a high-repetition rate excimer laser was used with a wavelength of either 248 nm (KrF, 1 mJ/cm^2, 100 Hz) or 351 nm (XeF, 0.2 mJ/cm^2, 500 Hz).

For the liquid-phase oxidation experiments, the sample was placed in a quartz cell containing 20 MΩ deionized water. Before each experiment, N$_2$ was bubbled through the water to remove any residual oxygen. The gas-phase oxidation was performed either in laboratory air or in the preparation chamber of an UHV system with a base pressure of 10^{-8} Torr.

The gallium arsenide samples used in this experiment were n-type of (100) orientation, doped with Si to $n = 10^{18}$ cm^{-3}. A comparison of oxidation rates was made using p-type GaAs doped with Zn to $p = 10^{18}$ cm^{-3}. Before oxidation experiments, the semiconductor surface was cleaned with trichloroethylene, acetone, methanol, deionized water, and 50% NH$_4$OH aqueous solution. The sample was then dried with N$_2$. In addition, samples used in the UHV system were subsequently heated to 540 °C, treated with mild argon-ion sputtering (500 eV, 1 μA/cm^2, 10 min) to remove residual oxide and carbon from the surface, and then annealed at 540 °C for 5 min.

The thickness of the oxide was determined in an UHV multiprobe system described previously.[9] The x-ray photoelectron spectroscopy (XPS) data was taken using Al $K\alpha$ rays (1486.6 eV) with a background pressure of $\sim 10^{-10}$ Torr. An estimate of oxide thickness is made by using the area ratio of the shifted oxide peak and the substrate peak of the same core level,[10] either As $3d$ or As $2p_{3/2}$, depending on the oxide thickness. The mean escape depths for electrons in GaAs have been determined to be 25 Å for the As $3d$ level, and 7 Å for the As $2p_{3/2}$ level.[11] This allows us to determine oxide thickness in the range of 1–50 Å. Auger electron spectroscopy (AES) depth profiling has also been used as a relative measure of oxide thickness.[9]

IV. LIQUID-PHASE OXIDATION

There is a considerable amount of work reported in the literature concerning reaction products on GaAs surfaces formed in an aqueous ambient.[12,13] An important reason for

studying the liquid-phase oxidation of GaAs is that semiconductor dissolution always involves an oxidation state change.[14] Typically, the oxide growth is controlled by applying an external bias, i.e., anodic oxide growth.

In our experiments, however, the room-temperature oxide growth in deionized water was obtained by illuminating the semiconductor surface with cw laser sources radiating at different wavelengths. The measured oxidation rates for GaAs have been shown to exhibit a strong wavelength dependence. The oxidation is much more rapid in deep-ultraviolet than in near-ultraviolet or visible wavelength regions. Detailed x-ray photoelectron spectroscopy (XPS) and Auger analysis of the oxide layer produced in our liquid-phase experiments has been reported elsewhere,[9] thus, only a brief description is given here. Independent of the laser wavelength used, compositional analysis of the oxides showed the almost complete depletion of As. The grown overlayer was a Ga-rich oxide. This is consistent with Pourbaix diagrams[15] on Ga and As, i.e., with the known relative solubility of Ga and As oxides. Deionized water can oxidize GaAs, but it will also attack the photon-enhanced oxide to dissolve some of the oxidation reaction products. For example, As$_2$O$_3$ being highly soluble in aqueous solutions of all pH values, will dissolve in water at pH $= 7$. Ga$_2$O$_3$, however, is insoluble in water and thus remains on the surface. This known dependence of the solubilities of As$_2$O$_3$ and Ga$_2$O$_3$ on solution pH accounts for the chemical composition of the grown overlayers. Finally, it is important to point out that the nonstoichiometry of oxides formed in a liquid-phase has been used to alter the formation of metal–semiconductor interfaces[16] and is therefore a subject of on-going interest.

Due to the thickness of the oxide layer grown in the liquid phase, AES depth profiles were used to determine the wavelength dependence of the oxide growth in water. Visible light illumination did not produce a significant change in the oxidation kinetics, i.e., the measured oxide thickness was comparable with a typical native oxide thickness value. This behavior contrasts sharply with that observed for irradiation with ultraviolet light. In the latter case, the measured oxide thickness was much larger. For example, Fig. 2 shows the data for a 250-Å-thick oxide, grown after 20-min illumination with 257-nm-laser light in deionized water. The corresponding thickness for an oxide layer grown in 514 nm light
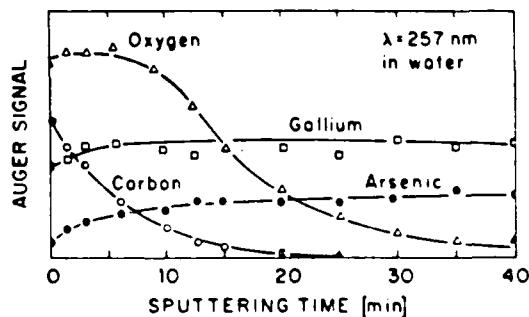
FIG. 2. AES depth profile of a GaAs sample immersed in water and exposed to 257 nm laser irradiation.

and that grown without laser irradiation was 80 and 70 Å, respectively.

The observed rapid, ultraviolet-induced oxidation in deionized water requires injection of holes from GaAs to the H_2O/O_2 redox level.[14] This redox level is centered around 0.4 eV below the top of the GaAs-valence band.[17] Thermalized holes in the valence band are not able to access this redox level efficiently and activate the rapid reaction; thus hot, nonthermalized holes are required in GaAs–water interfaces to obtain the observed reaction rates. Some previous experiments have shown that hot electrons are active in promoting surface chemical reactions.[18] More recently, Boudreaux *et al.*[19] have discussed the possibility that holes or electrons reach the semiconductor–surface interface as hot carriers and can in this way initiate reactions that would need more energy than that available at the band edge. It is reasonable, therefore, that the hot holes excited in GaAs cause the oxidation reaction observed in liquid-phase oxidation.

V. GAS-PHASE OXIDATION

Wavelength-dependent enhancement of the GaAs oxidation reaction has also been observed in a series of experiments performed at room temperature in gaseous ambients. like the liquid-phase oxidation, the gas-phase oxidation is found to be more strongly enhanced by ultraviolet light than by visible light. In the gas phase experiments performed in laboratory air, the samples were chemically cleaned, irradiated with either cw or pulsed sources, and inserted into the UHV system for analysis. The primary method of determining the thickness of the grown oxide was the comparison of XPS peak areas. In contrast to liquid-phase oxidation, XPS showed that the oxide layer grown in gas phase was stoichiometric with regard to arsenic and gallium, as has been generally reported.[20-22] The peak area ratio of the As $2p_{3/2}$ to the Ga $2p_{3/2}$ for a clean surface was ~ 1.4. The ratio for the oxide layer was also found to be ~ 1.4. This ratio has been found to correspond to the photoemission of a stoichiometric GaAs surface grown by MBE.[23]

A comparison was made of n-type GaAs oxidation and p-type GaAs oxidation under similar conditions with different wavelength irradiation in air. A typical result is when 257-nm cw laser irradiation was used for 20 min in moist air. The oxide thickness on n-type substrate and p-type substrate was 21 and 20 Å, respectively. Our results agree with previously reported studies where no difference was seen between the p-type and the n-type oxidation rates.[20-22] However, previous studies have not reported the effect of deep-ultraviolet light on the oxidation of GaAs.

The relation of wavelength of irradiation to oxide thickness for fixed exposure time was initially investigated in laboratory air. A summary of results is shown in Fig. 3, where weak enhancement is seen for wavelengths of 351 nm and greater, while a strong enhancement is seen for the deep ultraviolet wavelengths of 257 and 248 nm. The open circle shown at zero photon energy is the oxide thickness from a control sample. This sample was cleaned and handled in a manner similar to the other points in Fig. 3, but it was not exposed to laser radiation. It shows the thickness of a typical
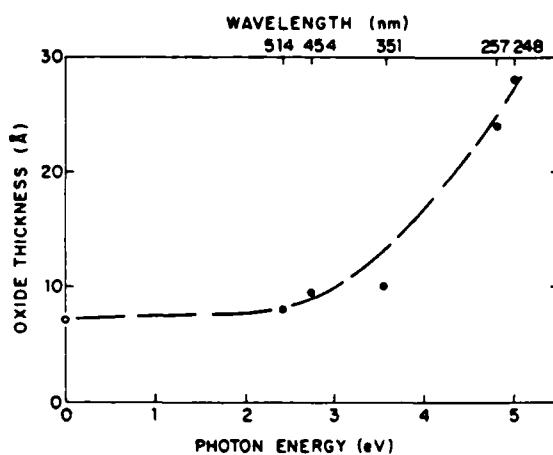

FIG. 3. GaAs-oxide-layer thickness determined by XPS for different wavelength laser irradiation at 50 mW/cm². Each irradiation was performed in laboratory air for 20 minutes.

residual oxide formed during sample handling in air. The enhanced oxidation rates can be clearly seen on top of this residual thickness. It is interesting to note that similar results were found regardless of whether the source was cw or pulsed. In either case, the laser intensities were low so as not to cause thermal oxidation.

The wavelength dependence of the oxidation of GaAs was confirmed by a comparison of oxides grown on clean surfaces in controlled ambients. A chemically cleaned GaAs sample was loaded into the UHV system and then sputtered and annealed as described. XPS was used to confirm that no oxide was on the surface within detection limits. The clean sample was transferred to a preparation chamber into which gases were controllably leaked. Laser light was directed at the sample surface through a Suprasil window on the preparation chamber, and the sample was then transferred for analysis without exposure to laboratory air. Using an environment of 10-Torr water vapor and 10-Torr O_2 in the preparation chamber, 20-min irradiation with an excimer laser at 351 nm (XeF) produced only 1.1 A of oxide. Under similar conditions, with 248 nm (KrF) illumination, an oxide layer of 8.1 Å formed. The relative difference in the shifted oxide peak can clearly be seen in the XPS spectra of Fig. 4 where a strong wavelength dependence for the oxidation of GaAs is confirmed.

To understand the growth characteristics of the GaAs oxide, additional experiments were performed in the well-controlled environment of the UHV system. The ambient in the preparation chamber was varied from 10 Torr to 1 atm of O_2, and from 0 to 10 Torr of water vapor. A typical exposure time was 20 min. When the GaAs sample was exposed to moist O_2 for 20 min without laser irradiation, no oxidation was detected. Also, when the sample was irradiated in 10 Torr of water vapor with no O_2 introduced, no oxide was detected.

A comparison was made of the oxidation rate in a moist O_2 ambient and a dry O_2 ambient. The moist ambient consisted of a mixture of 10-Torr O_2 and 10-Torr water vapor. The dry ambient was 10-Torr O_2 only. As shown in Fig. 5,
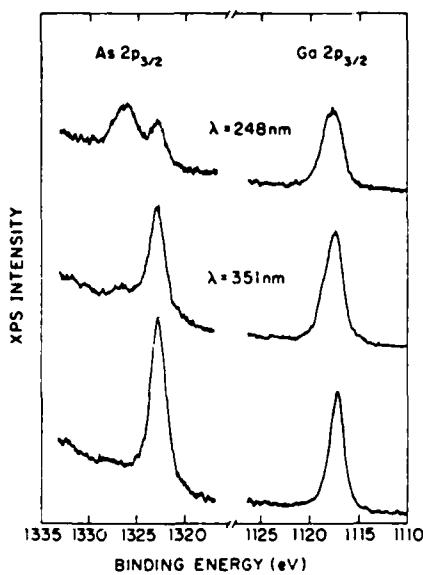
FIG. 4. X-ray photoelectron spectra of As $2p_{3/2}$ and Ga $2p_{3/2}$ peaks of the clean and laser-irradiated samples. Laser exposure was for 20 min in a 10 Torr O_2/10 Torr H_2O ambient.

for oxide layers up to ~ 5 Å, there is no detectable difference in the oxidation rate by illumination in moist O_2 and dry O_2. However, for thicker oxides, those irradiated in a moist O_2 ambient grow much faster than those irradiated in dry O_2. We believe that enhancement due to the moist environment is similar to the enhancement of water vapor in silicon oxidation where hydroxyl groups are incorporated in the oxide layer. This oxide is more porous than the oxide grown in dry O_2, allowing more rapid transport of reactants through the oxide layer.[24] Further experiments are needed to determine which reactants are being transported through the oxide and to confirm that the hydroxyl groups are incorporated.

Another series of observations is shown in Fig. 6. Here we have measured the O_2 exposure to the GaAs surface in terms of Langmuir, determined by the oxygen pressure and exposure time. The time of the oxygen exposure also equals the


FIG. 5. GaAs-oxide layer thickness as a function of laser exposure time. The results are compared for dry and moist O_2 ambient.
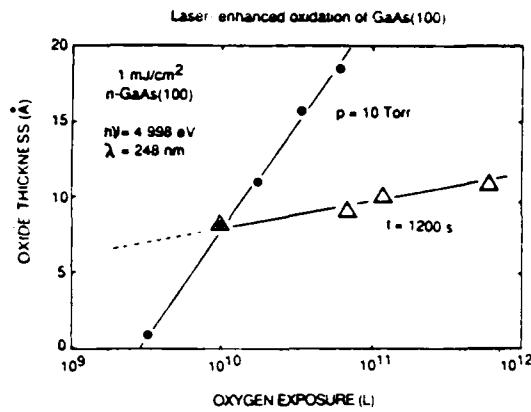

FIG. 6. GaAs-oxide layer thickness as a function of O_2 exposure (1 L = 10^{-6} Torr s). The results are compared for cases when either the O_2 pressure or the exposure time was held constant.

time of the laser exposure. The figure shows very different oxide thicknesses for the same O_2 exposure produced through different means. When the exposure time is held constant, and the O_2 pressure is increased, the oxide thickness only slowly increases. On the other hand, when the O_2 pressure is held constant, while the exposure time is increased, the oxide thickness increases much more rapidly. It has been suggested that a more appropriate measure of exposure would include a dependence on the number of carriers generated at the surface.[21]

VI. DISCUSSION AND CONCLUSIONS

Several groups have postulated that photogeneration of carriers is responsible for the enhanced oxidation of GaAs under illumination,[21,25,26] but a detailed mechanism of the role of carriers is not yet developed. For gas-phase oxidation it has been suggested that the attachment of hot electrons to physisorbed O_2 molecules, with the resultant lowering of dissociation energy, is responsible for the faster oxidation rate.[27] A similar effect has been claimed for the photoenhanced oxidation of silicon.[28] Another effect that has been suggested is the surface recombination of photogenerated electron-hole pairs which provide energy to accelerate the oxidation reaction.[22] However, the strong wavelength dependence of the photoenhancement of GaAs oxidation, in either gas or liquid-phase experiments has not been previously reported.

In order to attribute the wavelength dependence of oxide growth to the effect of photogenerated carriers, we need to eliminate several other possible effects of illumination that are wavelength dependent. One effect is the direct photon excitation of O_2 molecules before they reach the surface. The dissociation limit for O_2 is known to be 5.1 eV.[29] Except where noted, this energy is greater than photon energies used in our experiments. In order to check that O_2 was not being excited by radiation in the gas phase, a cylindrical lens (1 m focal length) was used to focus a 248-nm (KrF)-excimer beam parallel to the surface of a clean-GaAs sample in a typical environment in the preparation chamber. The energy of these photons is 4.99 eV. For a 20 min exposure, virtually
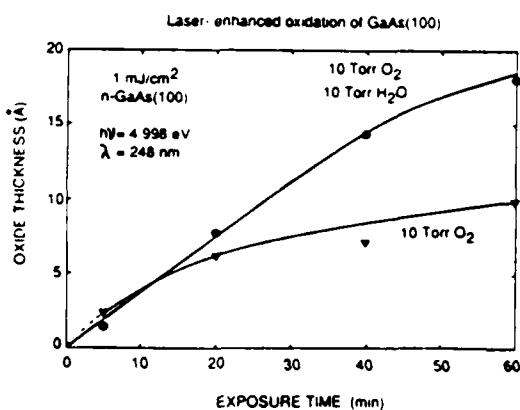
no oxide was formed. This is compared to 8 Å of oxide for a broad beam of 248-nm illumination normal to the surface. In addition, experiments were performed with 193 nm (ArF) laser light. The photon energy, $hv \sim 6.4$ eV, is now above the O_2 dissociation limit.[29] Consequently, a parallel illumination grew 25 Å of oxide, due to the known effect of ozone oxidation.[30] Direct beam illumination under similar conditions produced more than 70 Å of oxide (our XPS thickness limit due to the escape depth of the As $3d$ peak).

The effect of temperature rise can also be eliminated as a cause of oxidation. We confirmed experimentally that the temperature required for our typical oxide growth rate was ~ 450 °C. Calculations of the temperature rise of our GaAs surfaces due to the very low intensity radiation, typically 1 mJ/cm^2, give an upper limit of a 3 °C transient temperature rise.[31] Furthermore, due to similar absorption depths of 351 and 248-nm-laser radiation in GaAs, equivalent intensities of illumination would produce a nearly identical surface temperature rise in each case. However, we have seen that for equivalent exposures, the 248 nm light enhances the oxidation much more than the 351 nm light. Therefore, thermal effects can be ruled out as the cause of the wavelength dependence of the photoenhanced oxidation of GaAs. We thus attribute the wavelength dependence of the enhancement of GaAs oxidation to the wavelength dependence of the photogeneration of carriers.

At this time, the exact role of electrons and holes in the oxidation of GaAs is not known. It has been suggested that a possible role of the photogenerated electrons is that a chemisorbed oxygen molecule takes up an electron, enhancing the breakup of the molecule. It has also been suggested that photogenerated holes weaken the bonds between the substrate atoms at the surface. These effects would stimulate the uptake of oxygen.[21]

In conclusion, it is important to note that the energy of the photogenerated electron is nearly constant over the wavelengths used here (see Fig. 1) making it unlikely that electrons are responsible for the wavelength dependent enhancement mechanism. In contrast, the energy of the photogenerated hole varies greatly over these wavelengths, indicating that hot holes may be responsible for the wavelength-dependent enhancement of the oxidation of GaAs observed.[3]

ACKNOWLEDGMENTS

[1]H. Gerischer, J. Vac. Sci. Technol. **15**, 1422 (1978)
[2]F. A. Houle, J. Chem. Phys. **79**, 4237 (1983), **80**, 4851 (1984).
[3]D. V. Podlesnik, H. H. Gilgen, A. E. Willner, and R. M. Osgood, Jr., J. Opt. Soc. Am. B **3**, 775 (1986).
[4]H. R. Philipp and H. Ehrenreich, Phys. Rev. **129**, 1550 (1963).
[5]D. E. Aspnes, G. P. Schwartz, G. J. Gualtieri, A. A. Studna, and B. Schwartz, J. Electrochem. Soc. **128**, 590 (1981).
[6]M. Cardona, in *Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Beer (Academic, New York, 1966), Vol. 3, p. 125.
[7]J. R. Chelikowsky and M. L. Cohen, Phys. Rev. B **14**, 556 (1976).
[8]H. C. Casey, D. D. Sell, and K. W. Wecht, J. Appl. Phys. **46**, 250 (1975).
[9]C. F. Yu, D. V. Podlesnik, M. T. Schmidt, H. H. Gilgen, and R. M. Osgood, Jr., Chem. Phys. Lett. **130**, 301 (1986).
[10]T. A. Carlson and G. E. McGuire, J. Electron Spectrosc. **1**, 161 (1972/73).
[11]H. Gant and W. Mönch, Surf. Sci. **105**, 217 (1981).
[12]G. P. Schwartz, G. J. Gualtieri, J. E. Griffiths, and B. Schwartz, J. Electrochem. Soc. **128**, 410 (1981).
[13]B. Schwarz, CRC Crit. Rev. Solid State Sci. **5**, 609 (1975).
[14]H. C. Gatos and M. C. Lavine, in *Progress in Semiconductors* (Temple, London, 1965), Vol. 9, p. 132.
[15]M. Pourbaix, *Atlas of Electrochemical Equilibria in Aqueous Solutions* (Pergamon, New York, 1966).
[16]S. D. Offsey J. M. Woodall, A. C. Warren, P. D. Kirchner, T. I. Chappell, and G. D. Pettit, Appl. Phys. Lett. **48**, 475 (1986).
[17]W. H. Latimer, *Oxidation Potentials* (Prentice-Hall, Englewood Cliffs, NJ, 1964).
[18]M. J. Madou, K. W. Frese, Jr., and S. R. Morrison, J. Electrochem. Soc. **126**, 1827 (1979).
[19]D. S. Boudreaux, F. Williams, and A. J. Nozik, J. Appl. Phys. **51**, 2158 (1980).
[20]V. M. Bermudez, J. Appl. Phys. **54**, 6795 (1983).
[21]F. Bartels and W. Mönch, Surf. Sci. **143**, 315 (1984)
[22]K. A. Bertness, C. E. McCants, T. T. Chiang, P. H. Mahouwald, A. K. Wahi, T. Kendelewicz, I. Lindan, and W. E. Spicer, in *Photon Beam and Plasma Stimulated Chemical Processes at Surfaces*, edited by V. M. Donnelly, I. P. Herman, and M. Hirose (North-Holland, Amsterdam, 1987) (in press).
[23]J. M. Woodall, P. Oelhafen, T. N. Jackson, J. L. Freeouf, and G. D. Pettit, J. Vac. Sci. Technol. B **1**, 795 (1983)
[24]S. K. Gandhi, *VLSI Fabrication Principles* (Wiley, New York, 1983), p. 376.
[25]S. A. Schafer and S. A. Lyon, J. Vac. Sci. Technol. **19**, 494 (1981)
[26]W. G. Petro, I. Hino, S. Eglash, I. Lindau, C. Y. Su, and W. E. Spicer, J. Vac. Sci. Technol. **21**, 405 (1982)
[27]W. Mönch, Surf. Sci. **168**, 577 (1986)
[28]E. M. Young, and W. A. Tiller, Appl. Phys. Lett. **50**, 46 (1987)
[29]G. Herzberg, *Spectra of Diatomic Molecules* (Van Nostrand, New York, 1950).
[30]S. Ingrey, W. M. Lau, and N. S. McIntyre, J. Vac. Sci. Technol. A **4**, 984 (1986).
[31]J. F. Ready, *Effects of High Power Laser Radiation* (Academic, New York, 1971)

꿈5

LASER-DESORPTION OF POLYMER IN A PLASMA REACTOR

W. Holber[*], D. Gaines, C.F. Yu,

and R.M. Osgood, Jr.

Microelectronics Sciences Laboratories

and

Columbia Radiation Laboratory

Columbia University

New York, NY 10027

[*]Present Address: IBM - Thomas J. Watson Research Center

Yorktown Heights, New York 10598

ABSTRACT

Low-power ultraviolet (350 nm) laser light has been used to inhibit polymer formation on silicon in a plasma reactor containing CF_4/H_2 reactants. The resultant increase in etch rate has been studied as a function of hydrogen partial pressure. Auger electron spectroscopy reveals a reduction in polymer formation in the area illuminated by the laser.

LIMITED DISTRIBUTION NOTICE

In previous work[1,2], we have described some of the effects of cw laser radiation focused onto the surface of silicon in a plasma reactor. Briefly, it was found that when laser light illuminated the surface of silicon in a CF_4/O_2 plasma, the etch rate was locally enhanced. At high laser powers, the enhancement was primarily due to thermal effects. At low laser powers, the effect was primarily nonthermal and was dependent on the doping level of the substrate and the wavelength of the laser light.

We have now shown that addition of a polymerizing agent to the discharge increases both the relative and absolute magnitudes of the laser enhancement. The laser light controls the etch rate by reducing polymer deposition on silicon. This laser-induced desorption may be used to change both the etch rate and the etch profile in a plasma reactor. The contrast provided under polymerizing conditions with low-intensity laser illumination greatly exceeds that obtained with a CF_4/O_2 plasma. The mechanism here, as in previous work[1,2], is attributable to the production of electron-hole pairs.

The experimental arrangement is similar to that described in the previous work[1,2]. The plasma etcher was operated with 40 mTorr of CF_4 (flow rate = 37 sccm) and with 0 to 30 mTorr of H_2. The radio frequency (rf) power was 30 watts -- corresponding to an rf power density of 0.05 W/cm^2. The laser source was a krypton-ion laser, line tunable to various wavelengths between 350 nm and 800 nm. The majority of the work reported in this paper was done with the multiline near-uv output, 350-357 nm. In addition, in the

present experiment, a differentially pumped mass spectrometer with a high-pressure sampling head is attached to the plasma etcher chamber. With the plasma etcher at a pressure of approximately 100 mTorr, the mass spectrometer, pumped by a 50-1/sec turbomolecular pump, has a base pressure of 3×10^{-7} Torr.

Typically, an unmasked sample of (100) silicon was clamped to the water-cooled baseplate -- previous work showed this to be an adequate method of ensuring good thermal contact to the baseplate. The sample size and doping were 1 cm^2 and n-type (1 Ω-cm), respectively. Sample masking allowed the etch or deposition depth to be measured after the run with a surface profilometer. For some of the runs, an aluminum-masked piece of silicon was used to check the dark etch (or deposition) rate.

Figure 1 shows the change in the dark etch rate of the silicon, as the partial pressure of H_2, added to the 40 mTorr CF_4, is increased from 0 to 30 mTorr (this change corresponds to 0 to 43% of the total pressure, respectively). The dark etch rate drops an order of magnitude over this range of H_2 concentration. If the H_2 concentration is raised above ~ 30 mTorr, deposition of a polymer layer, instead of etching, occurs.

The differentially pumped, residual gas analyzer was used to look at changes in gas-phase components, as a function of H_2 partial pressure. It was not possible to monitor fluorine concentration directly, since fluorine atoms are very reactive and very few pass through the orifice into the residual gas analyzer chamber without reacting via wall or gas-phase collisions. However, it was possible to monitor HF concentration instead.

Figure 2 illustrates the variation in HF signal as a function of H_2 partial pressure. The curve indicates that the production of HF peaks at about 20-mTorr H_2 partial pressure, then declines. Since HF production is an important mechanism for scavenging fluorine from the gas, this implies that as H_2 partial pressure is increased, the fluorine concentration falls, then rises again.

Samples coated with a brownish polymer film during processing were removed from the reactor and examined with x-ray photoelectron spectroscopy (XPS) under ultrahigh vacuum conditions. Figure 3 shows the XPS spectrum of the coated surface. The peak corresponding to the C 1s level appears to have a multiplet structure, with five partially resolved components indicated by dashed lines. From the high binding energy side, the components represent, in succession, trifluorocarbon (292.9 eV), difluorocarbon (290.9 eV), monofluorocarbon (287.9 eV) alkyloxyl carbon (286.6 eV), and alkyl/graphitic carbon (284.6 eV).[3] The comparable intensities of all components suggest an extensively fluorinated structure for the polymer film.

In the Si 2p level, the small unresolved broader peak (binding energy at 102 - 103 eV) corresponds to silicon oxides and possibly some fluorides as well, although polymer deposition was the predominate process for these samples. The oxides were probably formed in air before the samples were placed in the surface analysis vacuum chamber. The larger peak at 99.3 eV originates from the silicon substrate.

The results obtained in this set of experiments are consistent with a process in which the silicon etch rate is limited at lower

H_2 concentrations by lack of gas-phase fluorine atoms, and at higher H_2 concentrations by the formation of a polymer layer on the silicon surface, which can block reaction sites. The work done by several other researchers on the role of hydrogen in CF_4/H_2 plasmas supports this explanation. In particular, although the role of H_2 in polymerization is very dependent on system parameters such as total pressure, flow rate, electrode spacing, and RF power, it has been generally found[4,5,6,7] that adding H_2 to a CF_4 plasma reduces the silicon etch rate through a combination of scavenging fluorine from the gas and forming polymer on the silicon surface.

When unmasked silicon samples were illuminated with a weakly-focused 350-nm laser light, an enhanced local etch rate was observed. The laser spot size on the sample was typically ~1/4 mm in diameter. By patterning the laser beam so as to have light and dark regions on the substrate, the ratio of the laser-enhanced etch rate to the dark etch rate could be measured as a function of H_2 partial pressure. These data are displayed in Fig. 4. In this experiment, the laser power density was approximately 80 W/cm^2 in the illuminated areas. From 0 to 10 mTorr of H_2, the enhancement ratio increased slightly. However, at about 15-mTorr of H_2 pressure, there was a large, abrupt rise in the relative light-enhanced etch rate. After this point, the etch rate again rose slightly with increasing H_2 pressure. The laser has its greatest effect at the higher H_2 concentrations, where it is known from previous studies that there is an adequate supply of fluorine in the gas to allow more rapid etching to occur if the surface is not blocked by polymer deposits.

When the unfocused 350-nm laser output, at an intensity of

approximately 5 W/cm^2, was used to illuminate the surface of the unmasked silicon sample, formation of the polymer layer was inhibited at the point of impingement of the laser beam. When the laser beam was directed on the sample, an area that appeared visually clear of polymer film formation resulted. Auger electron spectra were taken both inside and outside of the illuminated region. Results show a reduction of at least fifty percent in carbon, oxygen, and fluorine within the illuminated region. This indicates that the laser exposure substantially reduces the amount of polymer formation on the silicon surface; this reduction allows the etch process to proceed at a more rapid rate.

In our previous experiments[1,2], it was found that laser-enhancement of the etch process depended on the wavelength of the laser source. This effect was interpreted to be a manifestation of photogenerated carrier-induced surface chemistry. With this in mind, the experiment of Figure 4 was repeated using the 647-nm output from the krypton-ion laser, with approximately the same power density. It was found that the slow rise in the ratio of light-to-dark etch rates occurs again. However, the sharp increase at 15-mTorr partial pressure H$_2$ was missing.

The above result may be understood on the basis of a mechanism based on carrier-induced desorption. It has been previously found in experiments involving both wet and dry laser-induced chemistry that carrier production with UV light results in an increase in desorption, or other surface reaction rates. This effect results from the fact that UV light is absorbed sufficiently, ~100A, close to the solid surface that bulk recombination effects do not reduce

the photogenerated carrier concentration at the surface[8]. On the other hand, for 647-nm light, the absorption depth is ~3μm, and the carrier flux to the surface is greatly reduced. Note that for both uv and red light with the given spot size, the sample heating at the surface will be essentially the same, although negligible, at these power densities. The absence of the sharp etch rate enhancement at 15 mTorr partial pressure when 647-nm light is used may therefore be explained on the basis of carrier induced desorption.

Two potential applications of this laser-induced process would be to increase the anisotropy of silicon etched in a plasma reactor, without increasing the degree of ion bombardment, and to permit maskless etching using a projected, patterned laser beam. With regard to the first application, since the laser light has been shown to increase the etch rate in this system, it might be expected to change the directionality of the etch as well. To examine this possibility, silicon masked with a 10-μm bar pattern was illuminated with a weakly focused (~100μm) laser beam, with 15 mTorr of H_2 added to 40 mTorr of CF_4 in the plasma reactor. Profiles were analyzed using an SEM. Initial results indicate that the light yields a profile which is not only deeper, but also has a much cleaner wall and bottom structure than in the unilluminated case.

In conclusion, we have demonstrated that low intensity, cw, uv light at 350 nm can be used to inhibit polymer formation on silicon in a plasma reactor operated with CF_4 and H_2. The dependence of the laser-enhanced etch rate, as a function of H_2 partial pressure,

is due to variations in both gas-phase fluorine concentration and polymer formation on the silicon surface. Work is currently underway to try to use this effect to obtain improved anisotropy in the plasma reactor, without the damage associated with increased ion flux and energy.

ACKNOWLEDGMENTS

REFERENCES

1. W. Holber, G. Reksten, and R.M. Osgood, Jr., Appl. Phys. Lett. 46, 201 (1985).

2. G. M. Reksten, W. Holber, and R.M. Osgood, Jr., Appl. Phys. Lett. 48, 551 (1986).

3. Lieng-Huang Lee, ed. Characterization of Metal and Polymer Surfaces, Academic Press, New York (1977), Vol. 2, p. 12.

4. L.M. Ephrath, J. Electrochem. Soc. 126, 1419 (1979).

5. R. d'Agostino, F. Cramarossa, V. Colaprico, and R. d'Etolle, J. Appl. Phys. 54, 1284 (1983).

6. J.W. Coburn and E. Kay, IBM J. Res. Develop. 23, 33 (1979).

7. S. Pang and S.R.J. Brueck, in Laser Diagnostics and Photochemical Processing for Semiconductor Devices, R.M. Osgood Jr., S.R.J.

Brueck, and H.R. Schlossberg, Eds., North Holland, New York (1983).

8. W. Holber, <u>Laser-Assisted Plasma Etching</u>, Thesis, Columbia University (1986); D.V. Podlesnik, private communication.
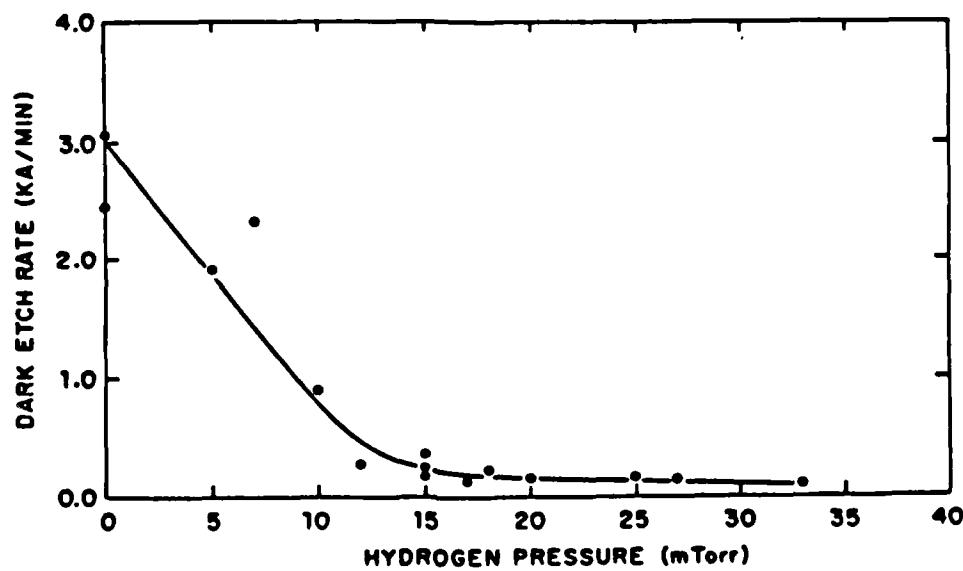
Figure 1:   Plasma etch rate of silicon vs. partial pressure of hydrogen added to 40 mTorr of CF_4.
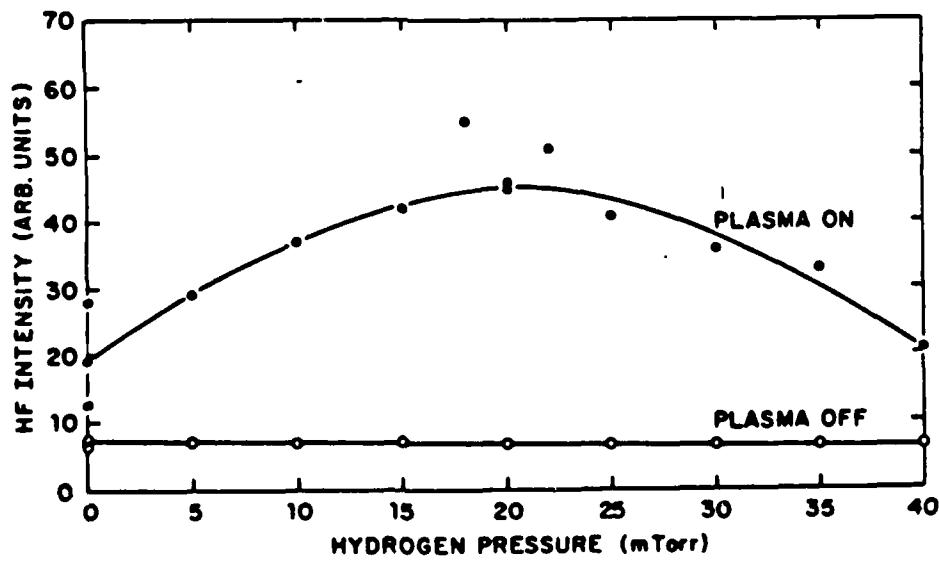
Figure 2:   HF (Mass 20) signal from differentially pumped mass spectrometer attached to etching chamber.
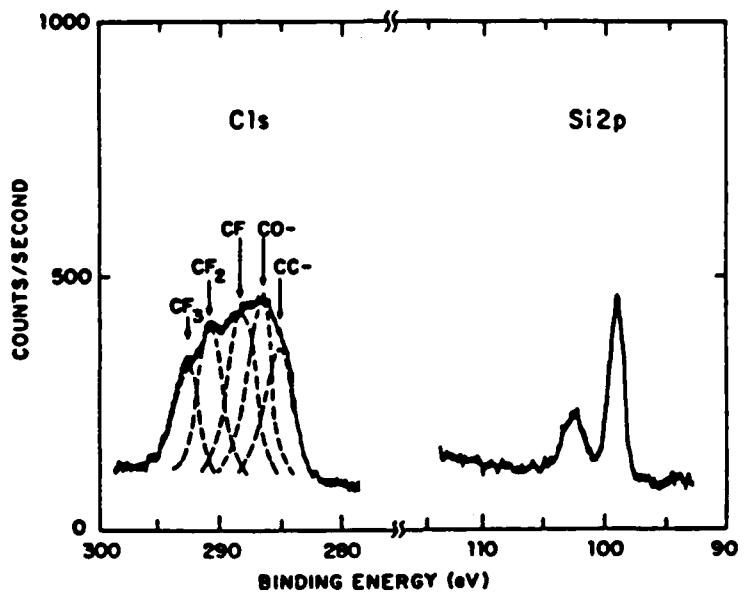
Figure 3:   X-ray photoelectron spectra of silicon surface with deposited polymer film.
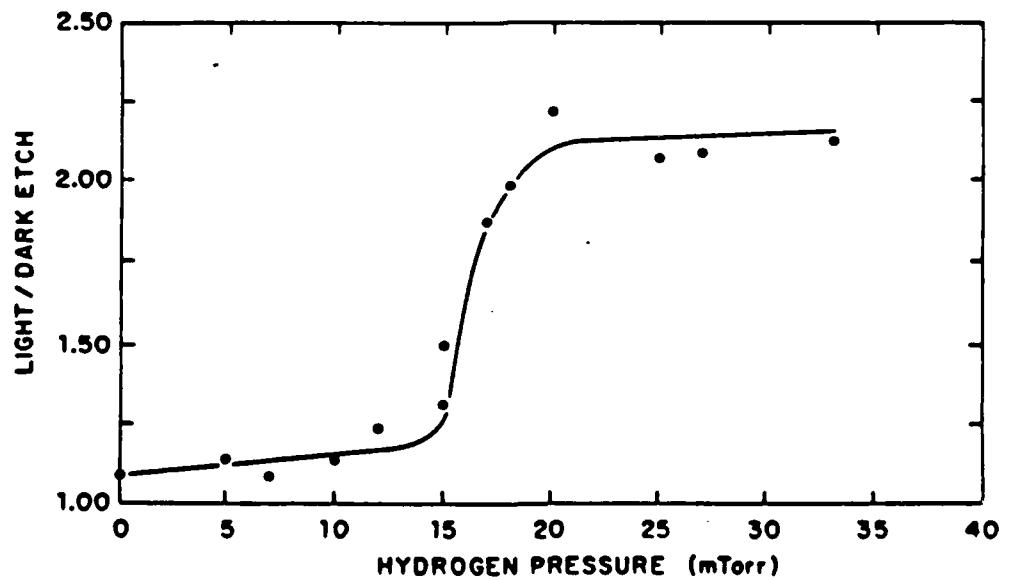
Figure 4:   Laser-induced etch enhancement vs. partial pressure of hydrogen added to 40 mTorr CF_4.

Copies may be requested from:

IBM Thomas J. Watson Research Center
Distribution Services 73-F11
Post Office Box 218
Yorktown Heights, New York 10598

Characterization of ultrathin SiO$_2$ films formed by direct low-energy ion-beam oxidation

C. F. Yu, S. S. Todorov, and E. R. Fossum

Columbia Radiation Laboratory and Department of Electrical Engineering, Columbia University, New York, New York 10027

Bombardment of silicon (100) surfaces at room temperature by an oxygen-containing low-energy ion beam is studied as an alternative to thermal oxidation to produce ultrathin oxide films. A self-limiting oxide thickness of about 50 Å is obtained by using ions with energy 100 eV or lower. Auger electron spectroscopy depth profiles of an ion-beam grown oxide and a thermally grown oxide show very similar composition. Grazing angle x-ray photoelectron spectroscopy indicates the presence of lower oxides of silicon near the surface. The capacitance–voltage characteristics of ion-beam grown oxides compare favorably with those of thermally grown oxides.

I. INTRODUCTION

Ultrathin gate oxide films are expected to gain increasing importance as metal–oxide semiconductor (MOS) devices are scaled to smaller dimensions. At the same time, low-temperature processes are already replacing many thermal fabrication steps in an attempt to minimize damage to the substrate. Many processes have been proposed as low-temperature alternatives to thermal oxidation of silicon, the most notable being plasma oxidation and chemical vapor deposition (CVD).[1-3] These methods, however, have not yet produced thin oxides suitable for MOS applications. We have studied the bombardment of (100) silicon surfaces by oxygen-containing ion beams and have successfully grown ultrathin device-quality oxides.[4]

The composition of the obtained ultrathin oxides is analyzed by Auger electron spectroscopy and x-ray photoelectron spectroscopy. The oxides are also characterized electrically by studying the capacitance–voltage (C–V) and current–voltage (I–V) characteristics of aluminum-gate MOS capacitors.

II. EXPERIMENTAL METHOD

Low-energy ion-beam oxidation uses a 2.5-cm-diam single-grid Kaufman-type source which produces a Gaussian ion beam with energies up to 100 eV. The maximum beam current density at the target is 150 μA/cm^2. The source uses argon and oxygen gases in varying ratios with an oxygen partial pressure in the low 10^{-4} Torr.

Ion-beam oxide MOS capacitors are fabricated on wafers which have a 5000-Å wet oxide grown at 950 °C and annealed for 30 min at the same temperature in dry N$_2$. The ion beam is then used to grow ultrathin gate oxides in windows etched in the thick oxide. The wafers are unheated; a thermocouple mounted on the substrate holder indicates typical temperature rises of < 5 °C above room temperature. Aluminum is evaporated and patterned to form the capacitors. After metallization the sample is annealed in forming gas at 400 °C for 5 min.

Control samples of MOS capacitors with thermally grown gate oxides are fabricated according to the same procedure. Here the gate oxides are grown in dry oxygen at 900 °C and annealed *in situ* in nitrogen.

Samples for surface analysis are prepared according to the same procedure without the aluminum deposition. Surface analysis was performed using a multiprobe surface spectroscopy system equipped with a concentric hemispherical energy analyzer. X-ray photoelectron spectroscopy (XPS) measurements were carried out with the Mg K_α line (1253.6 eV) at 240 W. The system was calibrated from the Au 4$f_{7/2}$ level with a binding energy of 83.8 eV. C 1s from residual carbon on the surface was verified to have binding energy 284.6 eV to eliminate any charging effects. Auger electron spectra were taken with a 3-keV electron beam with a 1-mm spot size on the surface. A differentially pumped ion gun was used to generate 3-keV argon ions to sputter the surfaces. The typical ion current density used was 1.2 μA/cm^2. The system pressure was kept below 5×10^{-10} Torr for XPS and 5×10^{-8} Torr for Auger electron spectroscopy (AES) and maintained at 2×10^{-7} Torr in argon while sputtering.

III. RESULTS AND DISCUSSION

Auger electron spectroscopy in conjunction with argon ion sputtering is used to study the depth profile of the oxide layer. As shown in Fig. 1, the O(KLL) and Si(LVV) Auger
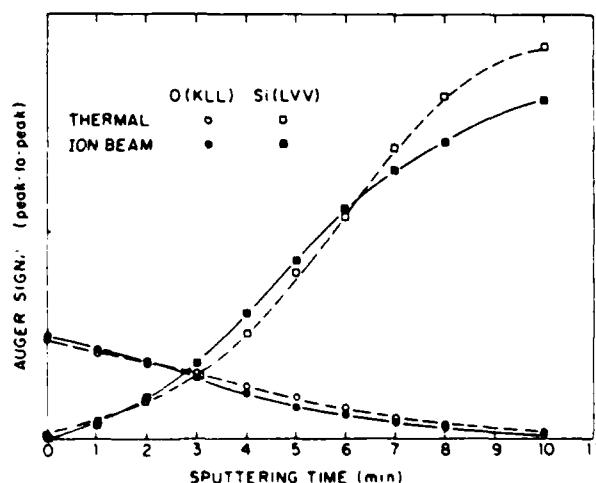

FIG. 1 AES depth profiles of thermally grown and ion-beam grown oxides

peak-to-peak signals are plotted as a function of sputtering time for both an ion-beam grown oxide and a thermally grown oxide about 50 Å thick as measured by ellipsometry. For the Si(LVV) emission only the signal from the substrate is shown because of the greater sensitivity. The close resemblance of the two depth profiles indicates that the ion-beam oxide is of comparable composition to the thermally grown oxide, except that the latter has a somewhat sharper interface. Because of the limited depth resolution the actual interface for both samples is much sharper than it appears.

Angle-dependent x-ray photoelectron spectra of the ion-beam grown oxide are shown in Fig. 2 for increasing detector angles. θ is the angle between the detector axis and the surface normal. At low detector angles the Si $2p$ (SiO₂) signal is peaked at 102.9 eV, a typical value observed for very thin SiO₂.[5] However, at $\theta = 80°$ where the electron escape depth is about 5 Å, the peak position is shifted to 102.1 eV. This indicates the presence of lower oxides in the top few layers due either to the preferential sputtering or to the low mobility of the oxygen atoms incorporated in the film. The O $1s$ peak has an unresolved component which grows in relative proportion toward higher detector angle and is shifted toward lower binding energy by about 1.2 eV. It has been shown that the O $1s$ peak due to lower oxides in ultrathin thermally grown SiO₂ does not shift more than 0.3 eV.[6] Therefore, this component is probably associated with the radiation damage caused by the ion bombardment. Large sample size was used to avoid the projection loss at grazing angle. The large decrease of intensity at $\theta = 80°$ is probably due to misalignment of the detector axis and the sample rotation axis. Both the O $1s$ and Si $2p$ (SiO₂) peaks from the thermally grown oxide (not shown here) appear to be sharper than the ones from the ion-beam oxide and their shapes do not change at higher detector angles.

Figure 3 compares the $C-V$ and $I-V$ characteristics of a
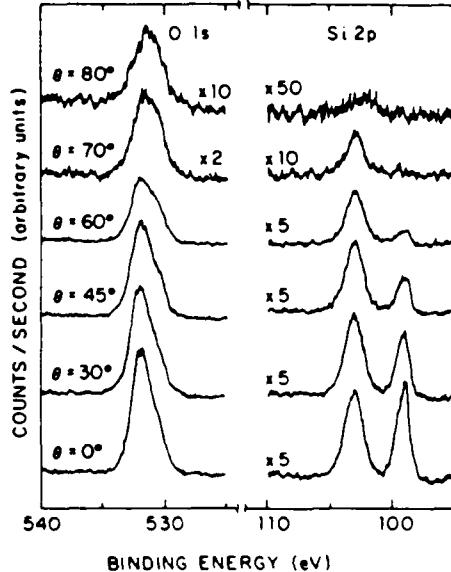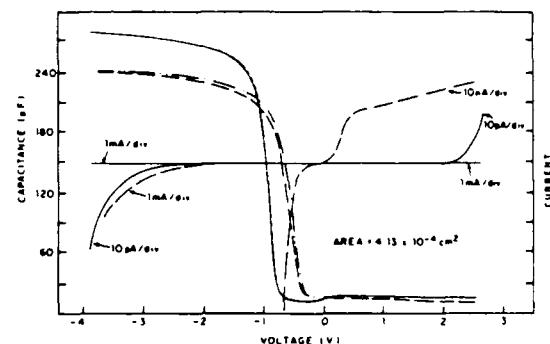
FIG. 3. Comparison of high-frequency capacitance–voltage and current–voltage characteristics for MOS capacitors with thermally grown (solid lines) and ion-beam grown (dashed lines) gate oxides.

capacitor with an ion-beam grown oxide and a thermally fabricated capacitor. The thickness of the ion-beam oxide is 56 Å; the thermal oxide is 52 Å thick. The ion-beam oxide is grown using a 100-eV ion beam with a 1⁻5 μA/cm² maximum current density at the target. The Ar:O₂ ratio in the source is 1:1. The sample is exposed for 6 min leading to a maximum dose of 2.8×10^{17} cm⁻².

The $I-V$ curves clearly show that the ion-beam oxide passes some current—in reverse bias its leakage current is three orders of magnitude greater than the thermal oxide leakage current. This may be due to a combination of higher thermal generation rates and a lower oxide impedance. Note, however, that the leakage current through the oxide does not hinder the proper functioning of MOS transistors fabricated with ion-beam gate oxides.[7] The leakage current is more than three orders of magnitude smaller than the drain–source saturation current.

A comparison of the $C-V$ curves, however, is much more favorable. Note that the characteristics are similarly steep in depletion and in the sharp onset of inversion. The small hysteresis of the ion-beam sample indicates the presence of some interface states. Quantification of the interface states is difficult due to the large leakage current. The threshold voltage shift of the ion-beam oxide is also indicative of uncompensated damage to the oxide or substrate. The behavior of the ion-beam oxide in inversion is easily understood from the $I-V$ curves—the deep depletion is due to some of the inversion layer charge leaking away. The observed rise in the $C-V$ characteristics in inversion is believed due to the coupling of minority carriers from surrounding inverted regions under the field oxide.

IV. SUMMARY

Low-energy ion-beam oxidation at room temperature has successfully produced ultrathin films of silicon dioxide suitable for use in MOS devices. The obtained film composition is comparable to that of thermally grown ultrathin oxide films.

ACKNOWLEDGMENTS

FIG. 2. Angle-dependent XPS spectra of ion-beam grown oxide. The Si $2p$ peak with lower binding energy (~99 eV), which disappears near grazing angle, corresponds to the substrate. The peak with higher binding energy (~103 eV) corresponds to the oxide.

[1]T. Roppel, D. K. Reinhard, and J. Asmussen, J. Vac. Sci. Technol. B 4, 295 (1986).

[2]S. I. Kimura, E. Murakami, T. Warabisako, H. Sunami, and T. Tokuyama, IEEE Electron Device Lett. 7, 38 (1986).

[3]A. K. Ray and A. Reisman, J. Electrochem. Soc. 128, 2424 (1981).

[4]S. S. Todorov, C. F. Yu, and E. R. Fossum, Vacuum 36, 929 (1986).

[5]P. J. Grunthaner, M. H. Hecht, F. J. Grunthaner, and N. M. Johnson, J. Appl. Phys. 61, 629 (1987).

[6]R. Flitsch and S. I. Raider, J. Vac. Sci. Technol. 12, 305 (1975).

[7]S. S. Todorov, S. L. Shillinger, and E. R. Fossum, IEEE Electron Device Lett. 7, 468 (1986).

LIGHT-GUIDED ETCHING FOR III-V SEMICONDUCTOR DEVICE FABRICATION

Dragan V. PODLESNIK

Microelectronics Sciences Laboratories and Center for Telecommunicatons Research, Columbia University, New York City New York 10027, U.S.A.

The rapid, ultraviolet-induced aqueous etching produces vertical, high aspect features in GaAs samples of different crystal orientations. Much of the speed and anisotropy of the etching is attributed to the formation of efficient hollow, optical waveguides. These guides have been characterized by measuring the optical loss and the field distribution within the guide. The optical loss is typically small and does not restrict the etching of deep features.

1 INTRODUCTION

Rate anisotropy, which results in a strong spatial directionality, is an important characteristic in many semi-conductor processing operations. In recent years, the work with the aniso-tropic etching has focused on its use for the machining of semiconductor materials. Vertical, high-aspect featu-res are an important requirement in the fabrication of advanced electronic and micromechanical devices[1]. The close spacings and small sizes of modern integrated circuits require vertical etching to eliminate the undercutting of adjacent structures. In addition,

the third dimension of a semiconductor wafer is currently being investigated as a means of providing additional electrical interconnections[2]. In particular, the interest in these through-wafer interconnections stems from three basic goals: to reduce the length of connections between devices for faster processing rates; to reduce the interference and crosstalk between interconnections; and to reduce the area occupied by interconnections on the surface of the semiconductor wafer. An example is a GaAs FET with via connections through the substrate[3]. Such transistor has higher gain at

microwave frequencies, higher power density, and potentially lower manufacturing costs than a conventional GaAs field effect transistor.

Finally, anisotropic etching has been extensively used in nonelectronic applications[1]. Even simple holes and grooves etched in a semiconductor wafer (typically silicon) can be utilized for many applications. One usage is the generation of high precision molds for microminiature structures. Patterns etched clear through the wafer can be applied in the area of ink jet printing technology. In particular, the hole on the bottom of the wafer is used as an orifice, typically about 20 um, for an ink jet stream[1].

2 MECHANISMS FOR ANISOTROPY IN ETCHING

Conventionally, etching anisotropy in single-crystal materials is achieved by relying on a crystal plane dependent process, such as reaction rates in wet chemical etching. This crystallographically sensitive etching has been successfully used to produce deep vertical features, but its utility is restricted by the requirement of specific and limited crystal orientations. Localized electrochemical jet etching has been also used to generate vertical holes in semiconductor wafers.

When etching occurs by an ion assisted reaction, etch rate anisotropy can be expected, because ions are incident normal to the wafer surface. When a semiconductor is plasma etched, incorporation of an appropriate gas additive results in the formation of a passivating film that prevents side wall etching. However, a perfectly vertical, through-wafer feature cannot be made with ion and electron beam sources because of the spreading of the particle beam and loss of the particle fluence with increasing feature depth. In addition, the massive particle bombardment produces typically an incurring damage of the semiconductor surface.

In this paper we will show that laser-induced aqueous etching with the ultraviolet beam can be used to make deep, high-aspect features irrespective of crystal orientation. If the laser beam is incident normal to the semiconductor surface, it could be expected that the light-assisted etching will occur primarily in the direction of the beam. But, unlike ion or electron beams, the laser beam is guided by the etched structure itself via glancing-angle reflections thus resulting in the vertical etching. This waveguiding effect coupled with the rapid etching at low laser intensities was first seen in the ultraviolet (257 nm) induced aqueous etching of GaAs[4].

the rapid etching at low laser intensities, was first seen in the ultraviolet induced aqueous etching of GaAs. Subsequent experiments have shown it to be a more general phenomenon, that is applicable to other semiconducting materials, e.g., Si and InP. Light-guided feature formation is observed when a focused, patterned, or scanned ultraviolet laser beam is used. In all cases in the ultraviolet, the etched structure confines and efficiently transmits the laser beam.

The possibility that the vertical, ultraviolet-induced etching relies on the semiconductor crystal anisotropy was ruled out after the etching experiments with GaAs samples of different crystallographic orientations. A weak crystallographic dependence is found, resulting in average etch rates of 7.9 9.11 um/min for the (111)A, (100), (110), and (111)B faces, respectively.

However, perfectly vertical features are obtained for all the etchings, proving that the crystal orientation does not affect the etched profiles. This characteristic is important since it allows maskless etching of vertical, high-aspect profiles in all crystallographic directions. Note that, whereas the ultraviolet anisotropic removal is virtually independent of crystallographic orientation of the semiconductor crystal, visible light does show a somewhat stronger crystallographic dependence, as reported in [8].

Figures 1 to 2 show typical profiles of microstructures such as via holes and slits formed through a GaAs wafer. The slits were etched by scanning the wafer perpendicular to the axis of the laser beam. In all experi-

Fig.1 SEM micrograph of slits etched through a GaAs wafer
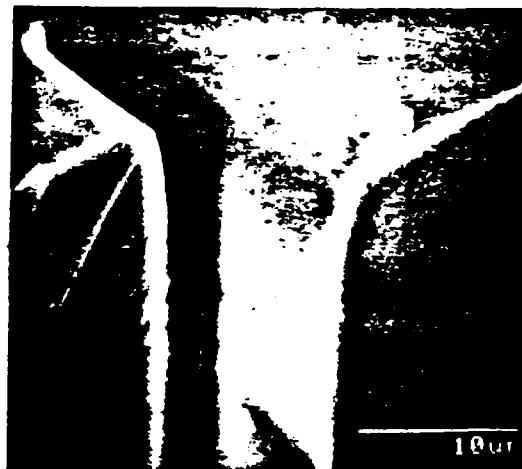
Fig.2 SEM micrograph of a etched groove in GaAs: close-up of the side wall.

ments, the laser beam was focused such that the beam waist was on the front of the semiconductor wafer. The entrance of the structures is well defined and the surrounding area is undisturbed. A distinct characteristic of these features is their smooth and vertical side walls. Their width, typically 2-3 um, remains virtually constant, independent of the etched depth. In contrast to this, the focused laser beam, with the 3 um waist diverges considerably over the corresponding distance, as shown in Fig. 3. The measured confocal beam parameter is only 25 um as compared, for example, to the perfectly vertical etch through the 200-um-thick wafer. This clearly shows that the processing beam is confined inside the hollow, etched structure which prevents the beam from diverging.

Because the vertical features are created by using a Gaussian beam, one might think that their profiles should also appear Gaussian. The vertical walls is a strong deviation from the Gaussian shape. In order to determine a mechanism for the formation of vertical waveguides, we monitored the development of holes, as shown in Fig. 4. Initially, the etch profile is essentially identical to that of the incident laser beam, as shown in Fig. 5. However, as the etch depth increases, the feature assumes a tubular, non-Gaussian profile, Fig 5.

The initial formation of these anisotropic features from a focused
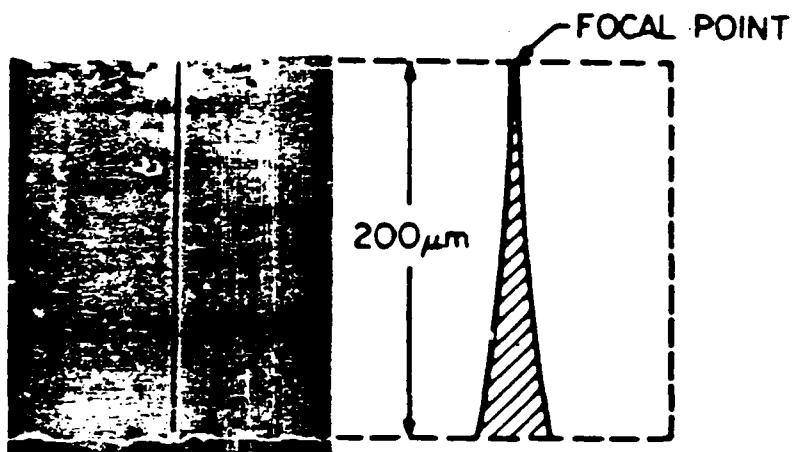
Fig.3 Comparison of the etched profile with free-space propagation of the focused laser beam.
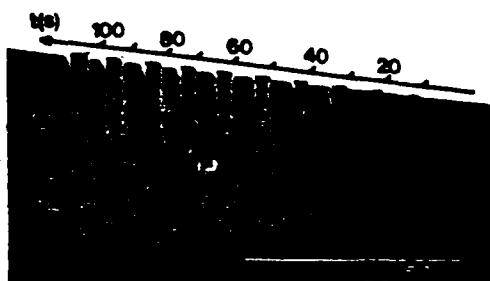
Fig.4 SEM micrograph of the temporal development of the vertical via hole.
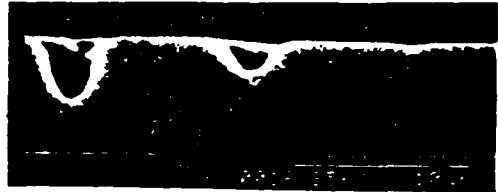
Fig.5 Initial, Gaussian-shape etchings and the formation of tubular, non Gaussian profiles.

laser beam with a beam waist of 4 um can be calculated with a model based on the use of ray optics. The model includes the very important angular dependence of reflectivity and absorption, which for the case of a lossy, metal like material must be obtained from a modified version of the Fresnel equations. In this case, the surface slope, which is equal to the local incident angle of the laser beam, is given by

$$\theta(z,r,t)=\tan^{-1}[dz(t)/dr(t)] \quad (1)$$

where $z(t)$ and $r(t)$ are, respectively, the time-dependent vertical and radial coordinates of the surface structure. The two components of the local interface velocity of removal are given by

$$(dz/dt)=[kI(r)(1-R(\theta))\cos(\theta)]\cos(\theta) \quad (2)$$
$$(dr/dt)=[kI(r)(1-R(\theta))\cos(\theta)]\sin(\theta) \quad (3)$$

where $I(r)$ represents the incident Gaussian beam and $(1 - R(\theta))$ is the local absorption. The interface velocity, in braces, is assumed to be linearly pro-

portional, by the reaction rate constant k, to the local absorbed light intensity. The $\cos(\theta)$ term in Eqs. (2) and (3) represents the change in incident power deposited on the surface due to surface tilt. The last term in Eqs. (2) and (3) represents the directionality of etching with this surface tilt. Material removal is always perpendicular to the plane of the surface.

During the very initial stages of formation of, say, a hole, with a Gaussian laser beam, all the angular dependent terms are inconsequential. As a result, initially the surface should follow the intensity distribution of the laser beam. Therefore the initial feature will be Gaussian.

As the feature walls become more vertical, the angular-dependent terms mentioned earlier become more important. Even then, for an incident Gaussian laser beam, the surface structure

takes a non-Gaussian shape. However, despite this non-Gaussian effect, the above model will not produce the striking features with vertical walls seen here. Many of these effects occur because light begins to be trapped inside and undergoes internal reflections as the hole deepens. Owing to the angle dependent reflectivity, there is still enough power reflected off 'the sidewalls to be subsequently absorbed as the light again strikes a different portion of the surface. This first internal reflection, or second strike, does contribute significantly to the formation of the features. Figure 6 shows this computer simulation of via hole development that includes internal reflections to the first order. These

Fig.6 Calculated hole development with (solid line) and without (dashed line) first internal reflection

reflections can have a pronounced impact on the developing shape of a cylindrical hole.

All the model figures display an overall appearance that is in agreement with an actual via. Eventually, a portion of the side wall becomes sloped to such a glancing angle that the laser light from the first strike is completely reflected. Also,when the hole is deep enough, the second strike resulting from the reflections off the bottom does not affect the shallower parts of the walls. Wall erosion therefore eventually stops. The incident beam is then optically guided through the hole, while vertical etching continues at the bottom. Under these conditions, etching can bee properly modeled as a self propagating waveguide.

In fact, much of the speed and anisotropy of the etching can be attributed to the properties of these optical waveguide-like features. In particular, the optical loss is small, even for 200- to 300-um-thick samples, and does not significantly limit the etch rate with increasing feature depth. Further, the continuation of etching even in the heavily solvated liquid at the bottom of a deeply etched feature is indicative of the active chemistry induced by the ultraviolet light.

5 OPTICAL WAVEGUIDE FORMATION

One of the most interesting consequences of the unusual laser chemistry described above is that it permits the formation of long cylindrical tubes or vias in semiconductor wafers. In fact, they are miniature, hollow semiconductor waveguides. This unusual microstructure has suggested a number of novel applications in microelectronics and integrated optics, including vertical, high-density optical fiber interconnects,[9] through-wafer optical light guides,[10] and high density, through-wafer vias for electrical interconnections from the front to the back of silicon or GaAs integrated circuits[3]

It was already suggested that the guiding of the laser beam is an important factor in determining via-hole profiles and etch rates The circular hollow dielectric and metallic waveguides have been thoroughly discussed by Marcatili and Schmeltzer[11] To illustrate some of the features of the hollow semiconductor waveguide, we applied a similar analysis for the via holes etched through GaAs samples The theory and the experiment showed excellent agreement Figure 7 shows the attenuation coefficient of the hollow GaAs waveguide as a function of hole diameter In agreement with the theo-
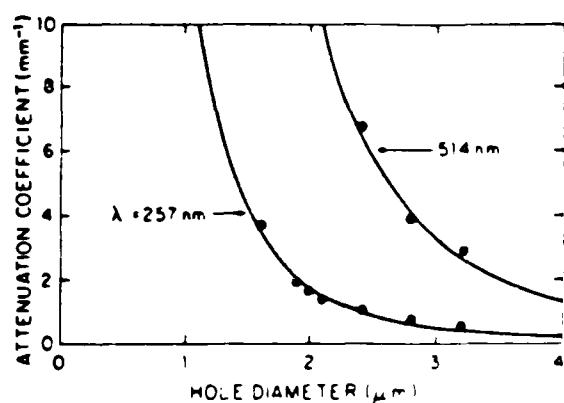
Fig 7 Attenuation coefficient of the hollow GaAs waveguide as a function of hole diameter, measured for 257 nm and 514-nm wavelengths

ry, the measured attenuation coefficient was 6.3 times larger for 514 nm than for 257 nm light Such high optical attenuation in conjunction with less efficient photochemistry is a reason that the visible laser light does not exhibit the same waveguiding effect

6 CONCLUSIONS

In this paper we have shown that the rapid, ultraviolet-induced process produces high-quality vertical etching through standard thickness GaAs wafers This directional etching is attributed in part to the formation of efficient optical guides, so that the laser processing-beam is guided by the etched structure itself Applications of this

technique to novel electro-optical
devices and through wafer electrical
connections are currently being
explored

At this point however many
important questions remain unanswered
For example, in the very deep etched
features, to what degree does capillary
flow caused by concentration or density
gradients influence the etching? Fur-
thermore in the format of submicro
meter ar par particular features mainly
it is known that surface electric
fields formed between dark and illumi-
nated areas significantly influence the
shape of the grating groove profiles
To what degree do surface fields influ-
ence the etched features seen in our
experiments? These questions as well
as the nature of the microscopic che-
mistry at the semiconductor surface
remain fertile ground for further
studies

ACKNOWLEDGEMENT

REFERENCES

[1] K E Petersen, Proc IEEE, 70, 420
(1982)

[2] T Morie, K. Minegheshi, and S.
Nakajima IEEE Electron Desvice Lett.,
EDL 24 411 (1983).

[3] L A D Asaro, J.V. DiLorenzo, and
R Fuki, IEEE Trans Electron Devices
EI 25 1218 (1978)

[4] D V Podlesnik, H H Gilgen, and
R M Osgood, Appl Phys Lett, 45, 563

[5] D V Podlesnik, T Cacouris, H H
Gilgen, and R M Osgood Jr, to be
published in Appl Phys Lett

[6] G Acre, M Rothschild, and D J
Ehrlich Appl Phys Lett, 46, 736
(1986)

[7] H E Philipp and H Ehrenreich,
Phys Rev 129, 1550 (1963)

[8] R M Osgood, A Sanchez-Rubio, D J
Ehrlich, and V Daneu, Appl Phys
Lett 40, 391 (1982)

[9] F F Frucnal, E R Fossum, and R.M
Osgood, Opt Lett, in press

[10] D V Podlesnik, H H Gilgen, and
R M Osgood, Appl Phys Lett 48, 323
(1986).

[11] E A.J. Marcatili and M A
Schmeltzer, Bell System Tech. J, 44,
1783 (1964).

ULTRAVIOLET-LIGHT-ENHANCED REACTION OF OXYGEN WITH GALLIUM ARSENIDE SURFACES

C.F. Yu, M.T. Schmidt, D.V. Podlesnik, E. S. Yang, and
R.M. Osgood, Jr.

Microelectronics Sciences Laboratories, Columbia University,
New York, NY 10027-6699

ABSTRACT

Light enhanced reaction of oxygen with gallium arsenide surfaces by irradiation with deep-uv, near-uv and visible light was studied using Auger electron spectroscopy and x-ray photoelectron spectroscopy for submonolayer and above monolayer regimes, respectively. The onset of a strong wavelength dependence of the enhanced oxidation was observed after the oxygen coverage reached more than half a monolayer. An abrupt threshold for this wavelength dependence was also observed at - 4.1 eV photon energy. Photodissociation of O_2^- (superoxide ion) formed by the photogenerated carriers can explain this strong wavelength dependence.

I. INTRODUCTION

Photon irradiation of semiconductors can nonthermally enhance chemical reactions on the surface. This effect has been generally attributed to the photogeneration of carriers at the semiconductor surface.[1] We have observed a strong enhancement of oxidation on gallium arsenide surfaces using deep-uv photons (248 nm, 257 nm), whereas only moderate enhancement was seen using near-uv to visible photons (350 nm to 514 nm) for experiments were carried out either in laboratory air ambient or in water solution.[2][3] This laser-enhanced oxidation can be used for low-temperature oxide growth and is potentially useful for device applications. Specifically, laser grown oxides have been used in our laboratory to increase the variation of Schottky barrier heights in metal-GaAs contacts.[4]

In this paper we report the results of a study of the light-enhanced oxygen reaction with GaAs surfaces both below and above monolayer coverage. Experiments were performed inside a stainless steel chamber on clean, well-defined surfaces irradiated with low-intensity excimer lasers. In other experiments we observed a wavelength threshold for the light-enhanced oxidation using different argon-ion laser lines under laboratory air ambient.

II. EXPERIMENTAL

2

A two-chamber ultrahigh vacuum system consisting of a surface analysis chamber and a sample preparation chamber separated by a differentially pumped load-lock seal was used in this study. Gallium arsenide surfaces of both (100) and (110) orientation doped with Si to $n = 10^{18}$ /cm^3 were used in the experiments. Samples were cleaned with trichloroethylene, acetone, methanol, deionized water, 50% NH$_4$OH, and then blown dry with nitrogen. For _in situ_ experiments, the sample was subsequently placed in the analysis chamber and subjected to mild argon ion bombardment (500 eV, 1 μA/cm^2) to remove the residual oxide, and then annealed at 540°C for 10 min. The sample was then moved into the preparation chamber filled with oxygen at a selected pressure. Through a quartz window a laser beam was introduced into the chamber and irradiated the sample for 1000 sec. After this process the sample was moved back to the analysis chamber for surface analysis. An excimer laser was used to irradiate the sample in the preparation chamber with either the KrF line at 248 nm (1 mJ/cm^2, 100 Hz) or the XeF line at 351 nm (0.2 mJ/cm^2, 500 Hz). At this low intensity the transient temperature rise on the surface is estimated to be below 10°C.[5] Some of the experiments were performed _ex situ_ with an argon-ion laser tuned to different wavelengths at a power density of 50 mW/cm^2. For these experiments samples were irradiated by the laser under laboratory air ambient after chemical cleaning.

3

X-ray photoelectron spectroscopy (XPS) was used to study the oxides formed on the surface. Al Kα (1486.6 eV) radiation was used to excite the $2p_{1/2}$ and 3d core-level emission for both gallium and arsenic. By making use of the large binding energy shift of the oxide arsenic atoms relative to the substrate arsenic atoms, the ratio of the oxide component to the substrate component can be obtained to estimate thicknesses between 1 Å to 50 Å as reported previously.[2] For submonolayer coverage (> 0.1 ML) of chemisorbed oxygen, Auger electron spectroscopy (AES) was used to study the oxygen coverage for its higher sensitivity. By taking the ratio of the Auger peak-to-peak intensities of O(KLL) to Ga(LMM) and using a layer model,[6] the fractional coverage of the oxygen on the surface can be determined up to a full monolayer. Care has been taken to perform the AES on a spot previously untouched by the electron beam after an ion-bombardment-annealing cycle to avoid electron beam effects.[7] The pressure of the analysis chamber was kept below 5×10^{-10} Torr for XPS and 5×10^{-9} Torr for AES.

III. RESULTS

The GaAs(110) surface was used here to study the enhanced oxidation by laser irradiation because it provides a uniquely defined surface after an ion-bombardment-annealing cycle.[6] Results of this study using XPS are shown in Fig. 1. Samples were cleaned by the ion-bombardment-annealing process

4

and then irradiated by laser light in the preparation chamber filled with 100 Torr dry oxygen. Only $As2p_{3/2}$ peaks are shown because the oxide and substrate components are well separated. The upper spectrum is of the surface irradiated with 248-nm light, the middle one with 351-nm light, and the lower one without any light. Without light irradiation only 1.9 Å oxide was formed (less than one monolayer). With irradiation of 351-nm light, 2.6 Å oxide was seen--a modest enhancement. With 248-nm irradiation, 9.6 Å oxide was grown under the same condition. This strong enhancement of oxidation by deep-uv light agrees with our previous results[2] conducted in laboratory air ambient on GaAs(100) surfaces.

Figure 2 shows the results of the AES study of the oxygen coverage on the GaAs(110) surface as a function of oxygen exposure with simultaneous irradiation of either 248-nm or 351-nm laser light, or without any light irradiation. At lower coverage the enhancement of the oxygen chemisorption is moderate and comparable for either light source within experimental uncertainties. However as the coverage approaches 0.7 monolayer, the enhancement of oxidation by the 248-nm light is seen to increase rapidly, while the enhancement by the 351-nm light remains moderate. This onset of strong wavelength dependence is also seen in the results obtained from the GaAs(100) surface as shown in Figure 3 in which the rapid increase of enhancement of 248-nm light occurs at 0.5 monolayer coverage. Also noted is that on

5

the (110) surface the oxygen coverage is smaller than that on the (100) surface for low oxygen exposure, indicating a lower initial sticking coefficient on the (110) surface.

Figure 4 shows the results of oxide growth as a function of photon energy using an argon-ion laser to irradiate GaAs(100) surfaces in laboratory air ambient for 20 min with each of the selected wavelengths. The dashed line at 7.5 $\overset{\circ}{A}$ represents the natural oxide grown on the surface in the course of the experiment as determined from the control samples. The oxide thicknesses due to light irradiation show a very sharp contrast for the oxidation enhancement. For photon energies above 4 eV the oxidation was about five times faster than for photon energies below 4 eV. Specifically, oxide growth for 275-nm and 257-nm argon-ion laser lines was much more strongly enhanced than that for 307-nm, 350-nm, 454-nm and 514-nm laser lines, respectively. This agrees very well with the results from the excimer laser study performed in the vacuum chamber, as shown in Fig. 1, where the 248-nm laser light shows a much stronger enhancement than the 351-nm light.

IV. DISCUSSION AND CONCLUSION

The sharp increase of oxidation enhancement at about 4 eV photon energy indicates a threshold phenomenon in the reaction. This threshold is close to the dissociation limit of O_2^- (superoxide ion), which has been studied

6

theoretically[8] and experimentally from the vibrational states[9] to be . 4.1 eV. O_2^- has been known to be a relatively stable ionic species in the gas phase from experiments of low-energy electron attachment to oxygen molecules[10]. The presence of O_2^- on the surface is thought to be the result of the attachment of photogenerated electrons to the physisorbed molecular oxygen on the surface and was suggested for the light-enhanced oxidation of both GaAs[11] and Si[12] surfaces. While the photodissociation of O_2^- is not known to have been performed experimentally, it has been suggested theoretically[8]. The photodissociation of O_2^- through the A - X transition is a much more likely event than that of the O_2 because of the selection rule restraint due to reflection symmetry is removed for the O_2^-. Dissociation of O_2^- will generate very reactive atomic oxygen species and thus cause a strong enhancement effect in the oxide growth.

The strong wavelength dependence of the light-enhanced oxidation, however, was not observed below about half a monolayer of oxygen coverage, as shown in Figs. 2 and 3, suggesting that a different mechanism may prevail in this regime. Bertness et al. have measured the activation energy of oxygen chemisorption on the GaAs(110) with and without light irradiation below monolayer coverage.[13] They found the activation energy to be 0.5 eV for dark oxidation and null under 514-nm light irradiation. It is thought, therefore, that under light irradiation (of above-band-gap photon

7

energy), physisorbed oxygens will become dissociatively chemisorbed immediately after the attachment of photogenerated electrons as long as there are available sites for chemisorption. This can explain the moderate and comparable enhancement of oxidation by both 248-nm and 351-nm photons at lower coverage because the photogenerated electrons are nearly constant in energy over this range of wavelength.[5]

Mönch has developed a simple model for the saturation coverage of oxygen on GaAs(110).[11] By assuming that dissociatively chemisorbed oxygen will be incorporated into two adjacent bonds in the surface Ga-As zig-zag chain, and that a repulsive interaction occurs between two pairs of chemisorbed oxygen along the chain, so they have to be separated at least by one unincorporated Ga-As bond, he arrived at a saturation coverage of ·56% of a monolayer. Beyond this coverage, further oxidation will occur by the diffusion of oxygen species through the first layer. We think that O_2^- would then form in the oxide by the attachment of ejected photogenerated electrons to the oxygen molecules. Photodissociation of this species will cause a much stronger enhancement of oxidation when above-threshold photons are used. While a direct experimental proof of the presence of O_2^- and its photodissociation is yet to be made, this mechanism offers a simple explanation for the strong wavelength dependence of the oxidation enhancement. The

8

results of this study suggest that the effect of the nonthermalized photogenerated carriers' seems to play a small role in the strong enhancement of oxidation by deep-uv photons as compared to the apparent photodissociation of O_2^-.

In summary, light-enhanced oxidation of GaAs has been studied with visible to deep-uv laser lines. The strong enhancement of oxidation by photons of energy higher than 4.1 eV can be explained as the result of the photodissociation of superoxide ion (O_2^-) formed by the attachment of photogenerated electrons to the oxygen molecules on or at the surface.

ACKNOWLEDGMENTS

FIGURE CAPTIONS

Figure 1. Comparison of the oxidation of GaAs(110) with or without laser irradiation. Shown are the x-ray photoelectron spectra of As$2p_{3/2}$ level.

Figure 2. AES study of the oxidation of GaAs(110) as a function of oxygen exposure with irradiation of 248-nm or 351-nm laser light or without any light irradiation. Data points beyond one monolayer are from XFS results.

Figure 3. AES study of the oxidation of GaAs(100) as a function of oxygen exposure with irradiation of 248-nm or 351-nm laser light or without any light irradiation. Data points beyond one monolayer are from XPS results.

Figure 4. Light enhanced oxide growth plotted as a function of irradiating photon energy. Dashed line represents the natural oxide grown on the surface in the laboratory air ambient.

REFERENCES

1. S.A. Schafer and S.A. Lyon, J. Vac. Sci. Technol. 19, 494 (1981); W.G. Petro, I. Hino, S. Eglash, I. Lindau, C.Y. Su and W.E. Spicer, J. Vac. Sci. Technol. 21, 405 (1982).

2. C.F. Yu, D.V. Podlesnik, M.T. Schmidt, H.H. Gilgen and R.M. Osgood, Jr., Chem. Phys. Lett. 130, 301 (1986).

3. C.F. Yu, M.T. Schmidt, D.V. Podlesnik and R.M. Osgood, Jr., J. Vac. Sci. Technol. B 5, 1087 (1987).

4. M.T. Schmidt, D.V. Podlesnik, H.L. Evans, C.F. Yu, E.S. Yang and R.M. Osgood, Jr., in this proceeding.

5. J.F. Ready, in Effect of High-power Laser Radiation (Academic Press, Orlando, 1971), p. 72.

6. W. Ranke and K. Jacobi, Prog. Surf. Sci. 10, 1 (1981).

7. W. Ranke and K. Jacobi, Surf. Sci. 47, 525 (1975).

8. G. Das, A.C. Wahl, W.T. Zemke and W.C. Stwalley, J. Chem. Phys. 68, 4252 (1978).

9. F. Linder and H. Schmidt, Z. Naturforsch. A 26, 1617 (1971); R.L. Gray, H.H. Haselton, D. Krause and E.A. Soltysik, Chem. Phys. Lett. 13, 51 (1972); D. Spence and G.J. Schulz, Phys. Rev. A 2, 1802 (1970).

10. L.M. Chanin, A.V. Phelps and M.A. Biondi, Phys. Rev. 128, 219 (1962).

11. W. Mönch, Surf. Sci. 168, 577 (1986).

12. E.M. Young and W.A. Tiller, Appl. Phys. Lett. 50, 46 (1987).

13. H. Okabe, in Photochemistry of Small Molecules (John Wiley & Sons, New York, 1978), p. 52.

14. K.A. Bertness, W.G. Petro, J.A. Silberman, D.J. Friedman and W.E. Spicer, J. Vac. Sci. Technol. A 3, 1464 (1985).
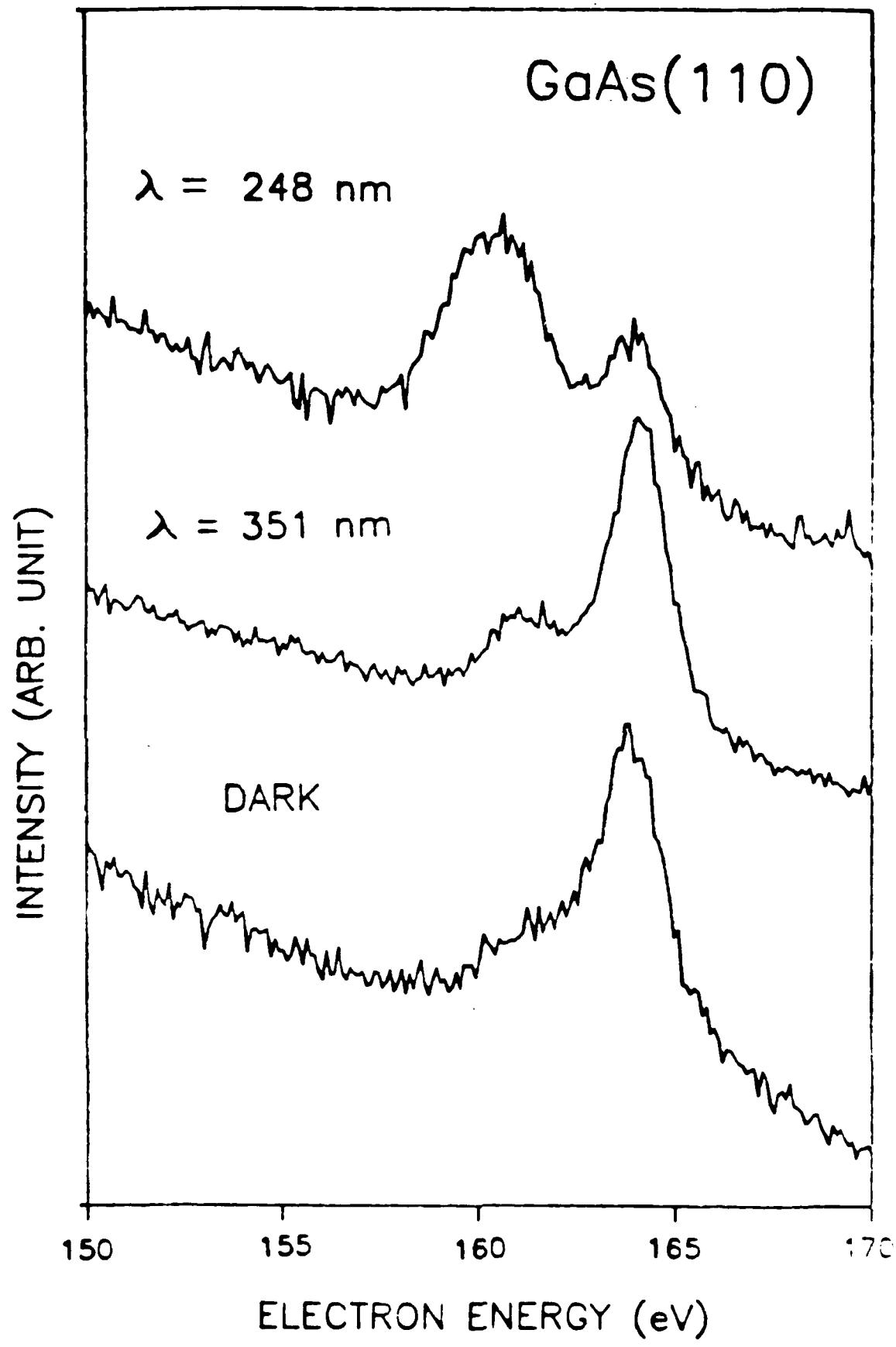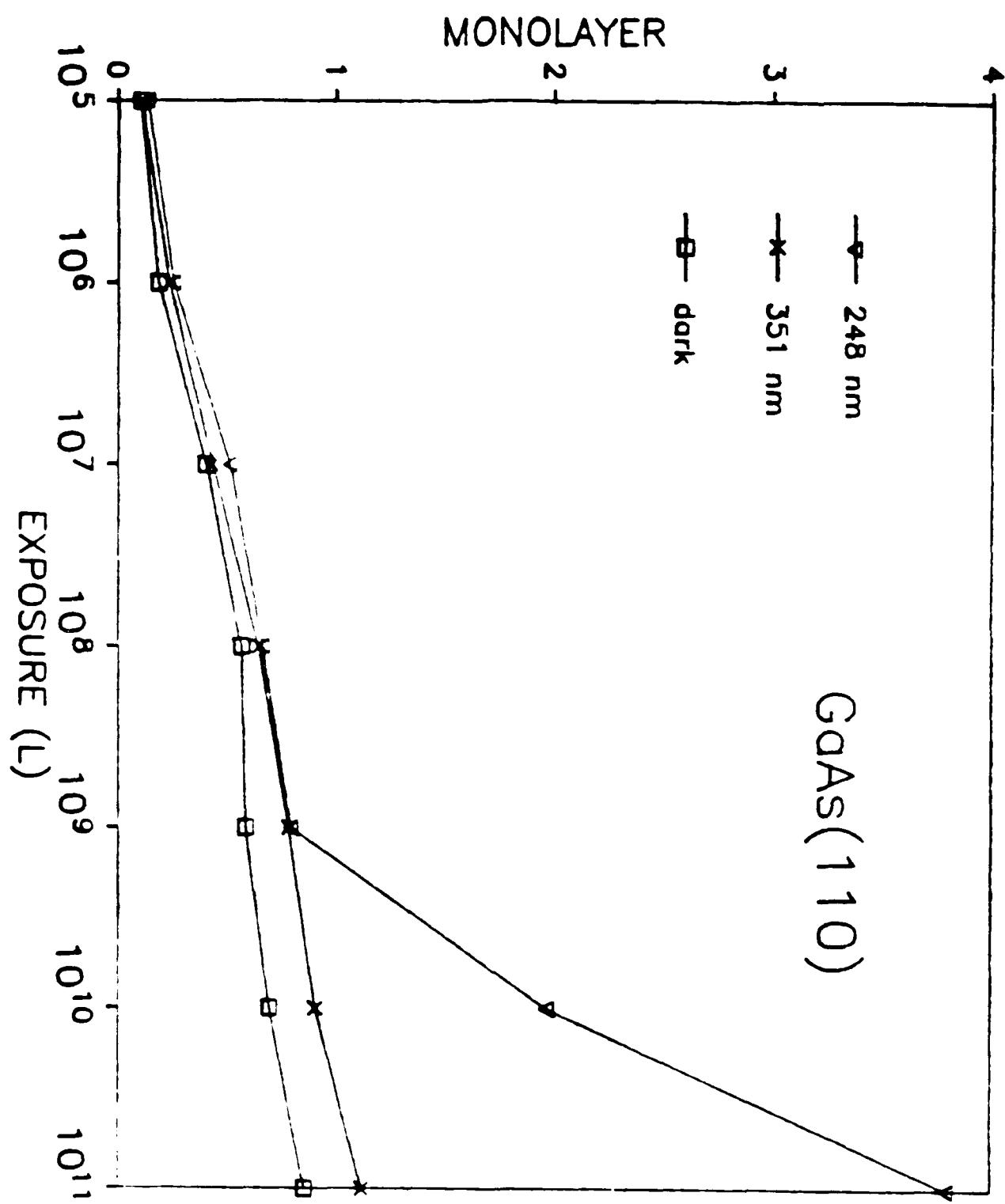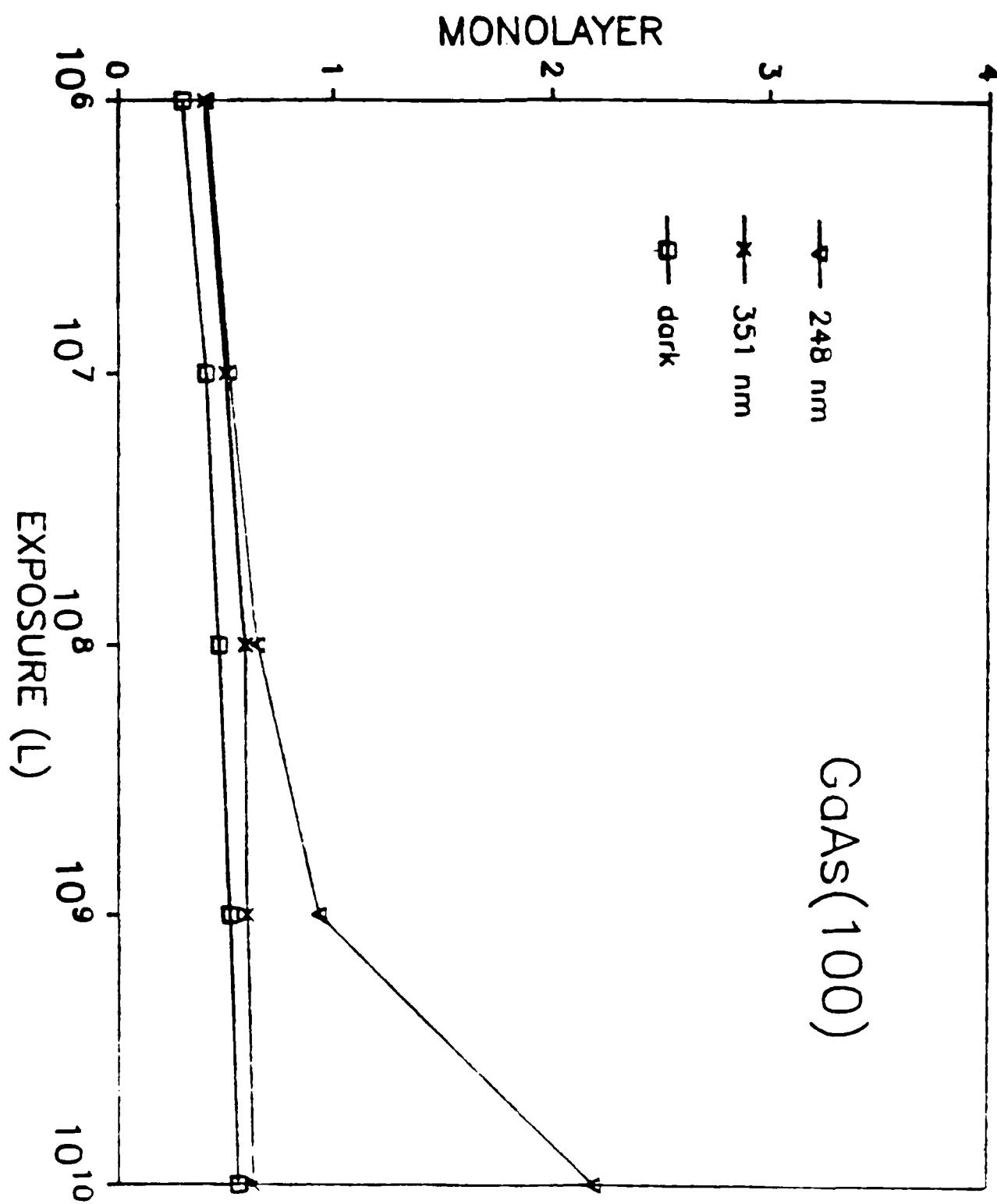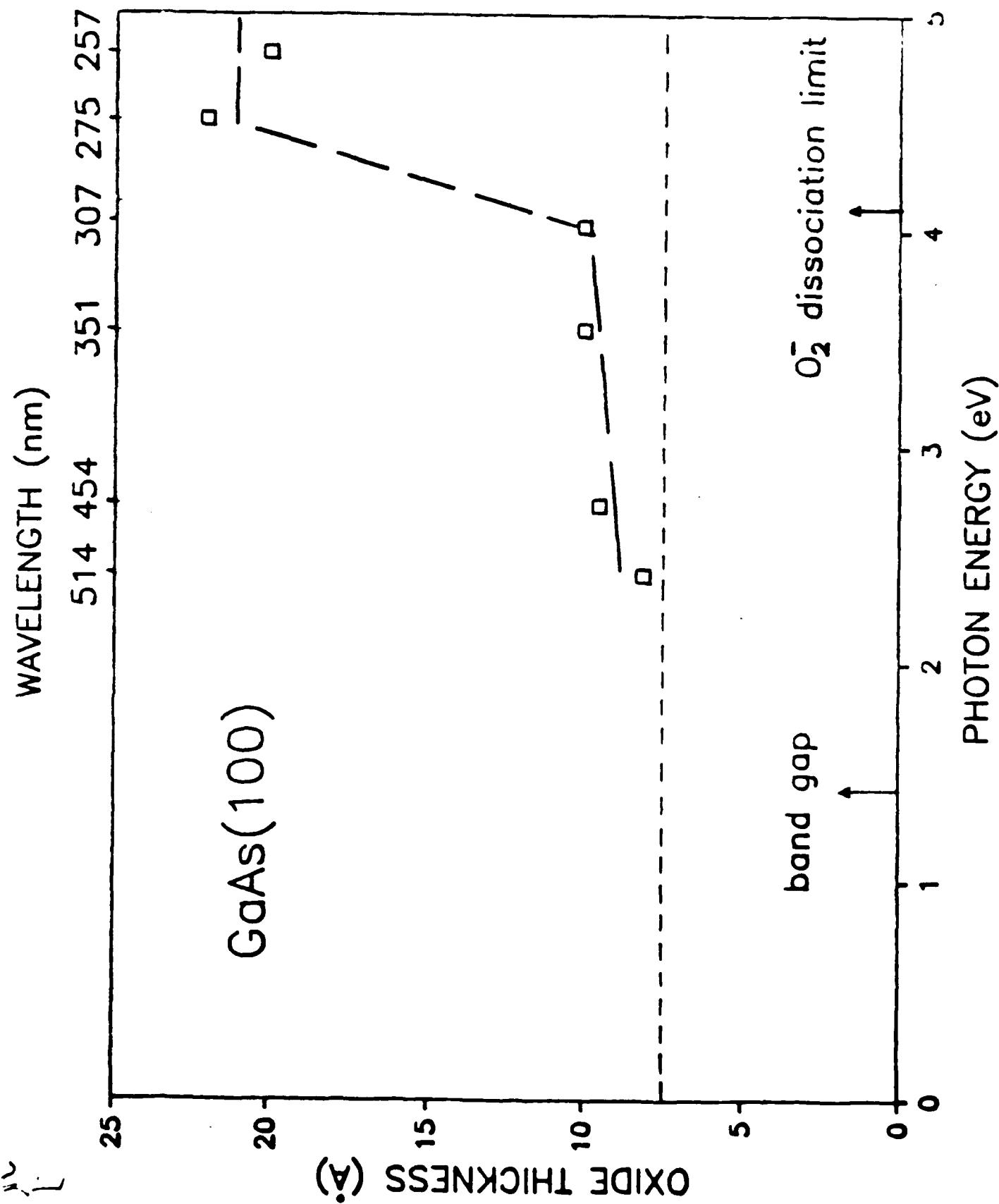
Fig. 2

GaAs(110)

Legend:
- 248 nm
- 351 nm
- dark

MONOLAYER (y-axis): 0, 1, 2, 3, 4

EXPOSURE (L) (x-axis): 10^5, 10^6, 10^7, 10^8, 10^9, 10^{10}, 10^{11}

WAVELENGTH (nm)

GaAs(100)

OXIDE THICKNESS (Å)

band gap

O_2^- dissociation limit

PHOTON ENERGY (eV)

AN IMPROVED DIFFERENTIAL VOLTAGE
TECHNIQUE FOR CAPACITANCE MEASUREMENT

X. Wu, H. L. Evans and E. S. Yang

Columbia University, New York, NY 10027, U.S.A.

Abstract—A differential voltage capacitance technique is described for measuring the capacitance of forward-biased Schottky diodes. This technique is based on the concept of an improved admittance bridge. Difficulties arising from the high conductive component were solved or minimized. Data were obtained for Ni–Si and Pd–GaAs Schottky diodes with quality factors as low as 0.001. Compared to accurate phase capacitance spectroscopy and other bridge methods, this technique is more reliable and easier to operate.

Admittance spectroscopy is a valuable method for investigation of semiconductor interface states. In the last twenty years, conductance and capacitance techniques have been widely used for studying MOS devices. However, when a similar method is applied to measure the capacitance of forward-biased metal–semiconductor contacts, large errors occur. This is because the capacitance signal of the diode is being overwhelmed by the thermionic emission conductance (G). This conductance is sometimes three or four orders of magnitude larger than the capacitance (ωC). Under this extreme condition, ordinary methods, such as the use of a Boonton capacitance meter, are no longer applicable. Even the accurate phase capacitance spectroscopy (APCS) method described previously[1] is limited by its instrumentation. First of all, the phase setting in APCS becomes extremely crucial. In practice, for a small phase error, $\Delta\phi$ (around 90°), we have:

$$\Delta(\omega C) = -G\,\Delta\phi, \qquad (1)$$

where $\Delta(\omega C)$ represents the resultant error of the measured susceptance and $\Delta\phi$ is in radians. For $G\,\omega C$ on the order of 10^4, even a 10^{-4} error in phase would cause a 100% error in the capacitance. Although the APCS method can set the phase to a very high accuracy (about 0.01°), it cannot ensure that there will be no phase shift during the measurement. Therefore, no matter how accurately the phase is set initially, a tiny phase drift will nullify the whole measurement. Furthermore, the sensitivity of the instruments, e.g. the lock-in amplifier (LIA), is limited by the conductance signal level. We know that the capacitance signal of a diode is very weak, especially at low frequencies. It is desirable to use a high sensitivity setting in order to get better noise performance. But this would overload the instruments because the capacitance signal is buried in a huge conductive component. The large in-phase signal frequently drives the LIA into a nonlinear region, making the measurement impossible. For this reason, we have implemented the following differential-

voltage scheme for measuring device capacitance when the quality factor ($\omega C/G$) is as low as 0.001.

The basic concept of the differential voltage capacitance spectroscopy (DVCS) is similar to that of a bridge circuit. Since the main problem arises from the high conductance that is inherently associated with the device, we propose to generate an in-phase signal to cancel out the conductive component of the diode, letting only the capacitive component be retained to pass through the phase sensitive detector. This way, the overload problem is avoided and the influence of the phase drift is also minimized. As a result the accuracy of the measurement is greatly improved. A similar bridge method was formerly suggested by Barret and Vapaille[2] and later modified by Greve[3]. Barret and Vapaille's method, using an equal-armed bridge, is tedious and time consuming since an adjustment of temperature for bridge balance is required at each data point. Greve's method, on the other hand, needs a floating a.c. signal source and a lock-in amplifier with both in-phase and quadrature outputs. The floating signal source readily picks up extra noise. Moreover, the biasing voltage in his bridge is applied to the diode through a current sensitive preamplifier, which is supposed for detecting the small a.c. signal. The non-separation of the a.c. and the d.c. paths results in an overload of the current preamplifier.

A schematic description of the differential voltage capacitance spectroscopy technique is shown in Fig. 1. A small a.c. reference signal, V_s, derived from the internal oscillator of a PAR 124 lock-in amplifier, is superimposed on a biasing voltage. This combined voltage is applied between point D and the ground. Two currents are generated. One, i_s, passes through the diode; the other, i_b, through a variable resistor, R. These currents are converted into voltage signals by resistors R', and are then fed into the inputs of the LIA through a differential preamplifier. The output of the differential preamplifier contains only the a.c. component of $V_a - V_b$, the d.c. portion is rejected. If R' is much smaller compared to $1/G$ and R, the
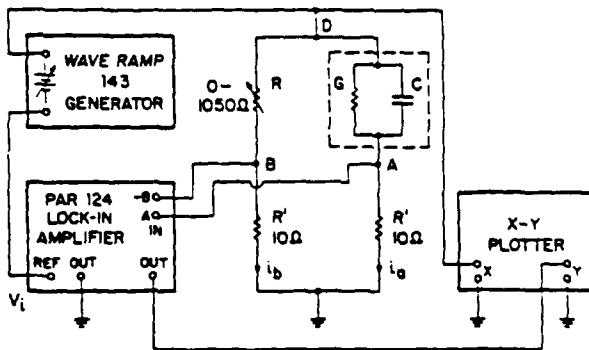
Fig. 1. Apparatus of the differential voltage capacitance spectroscopy.

output of the LIA is:

$$V_0(\phi) = V_i R'A[(G - 1/R)\cos \phi + \omega C \sin \phi], \quad (2)$$

where A is the voltage gain of the LIA and ϕ the phase setting of the phase sensitive detector. When the phase is set to $90°$ with a small error $\Delta\phi$, the output becomes:

$$V_0 = (90° + \Delta\phi) = V_i R'A[-(G - 1/R)\Delta\phi + \omega C], \quad (3)$$

and the error is:

$$\Delta(\omega C) = -(G - 1/R)\Delta\phi. \quad (4)$$

If the conductive component is nulled out, i.e. $G - 1/R \simeq 0$, the error will be very small even in the presence of a significant $\Delta\phi$. This can be contrasted with eqn. 1 where $\Delta(\omega C)$ is directly proportional to $G\Delta\phi$.

The measurement procedure is as follows. For each given frequency, the phase of the LIA has to be set at $90°$ first. This is accomplished by seeking a phase position of zero-output when V_b alone is being measured. The resistance of R may be chosen to be any value comparable to $1/G$ of the diode. While keeping the phase-range switch at $90°$, the phase vernier is varied until the output of the LIA is zero. Since V_b is in phase with the reference signal, the preceding adjustment ensures that the phase is set to measure only the capacitive component. The next step is to balance the conductive component. For this purpose, we turn the phase-range switch to $0°$ (without touching the phase vernier) and the input mode to $A - B$ so that the quantity $G - 1/R$ is examined. The variable resistor R is then adjusted until a zero output is reached. Turning the phase-range switch back to $90°$, the output of the LIA is now entirely due to the capacitance of the diode, i.e.

$$V_0 = V_i R'A\omega C. \quad (5)$$

The data are then collected by a plotter.

All the resistors used in the circuit should have small stray capacitance and inductance. The whole setup should be designed with care to keep the stray

capacitance as small as possible, especially at nodes A and B. In our experiment, for example, the sample holder, which was situated in a cryogenics coldhead to obtain various temperatures, formerly was connected to node A. Between the sample holder and the cold-plate there was a thin insulating sheet for thermal conduction. This configuration introduced an extra capacitance at node A in respect to the ground. A rational arrangement is to put the sample holder in connection with node D and let A connect to the probe needle. This reduces the stray capacitance significantly. The stray capacitance between node D and ground is less important, since it is parallel with the signal source.

One might use a voltage-controlled resistor or a photocell in place of the variable resistor to automate the balancing procedure[3]. However, it requires a lock-in amplifier with dual-phase output, and a few seconds delay at each data point to get a stable reading. A continuous voltage scan is difficult to achieve. Moreover, since the measurement is sensitive to stray phase shifts, it is desirable to keep all the circuit elements simple. In this circuit, a 1 kΩ and a 50 Ω, 10-turn potentiometer in series were used as coarse and fine adjustments. Cables and connectors were kept short to minimize parasitic inductance and capacitance.

The error caused by the experimental setup can be easily estimated by replacing the diode with a resistor corresponding to the highest conductance of the device. Under this situation, any susceptance measured is due to the setup. The stray capacitance was found to be less than 80 pf, which is about 1% of the measured device capacitance at low frequency (~ 10 nf at 100 Hz). The stray capacitance thus determined can be subtracted from the data if higher accuracy is desired. At high frequencies, the relative error may increase for the device capacitance gets smaller. Moreover, it was found that at frequencies above 30 kHz, the cable inductance became significant and our setup was not suitable for these measurements. The a.c. signal level used in the experiment should be kept smaller than kT/q. A large a.c.
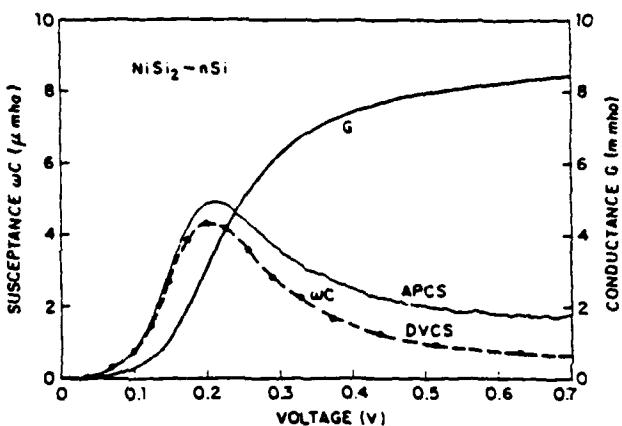
Fig. 2. The susceptance (ωC) and conductance (G) curves measured from a Ni–nSi Schottky diode at 108 Hz and 300 K. Note the scale of the conductance is 1000 times that of the capacitance.

signal will yield many high order harmonics because of the nonlinearity of diode I–V characteristics. However, a relatively large reference signal would be better for noise considerations. An optimum signal level depends on tradeoffs between the linearity and the measurement sensitivity. In our experiment best results were achieved around a 10 mV rms signal level, where both conditions seemed to be satisfied. The time constant of the LIA can be chosen according to the noise situation. An appropriate time constant (1 s in our case) should provide an adequate signal to noise ratio and, at the same time, an efficient measurement. This depends on the frequency: usually the lower the frequency the longer the time constant required.

The method was used to measure the forward-biased capacitance of a Ni–nSi diode[4,5]. Capacitance vs voltage data at various frequencies and temperatures were obtained. Figure 2 shows a typical result measured at 108 Hz and 300 K. The measure-

ment was repeated several times over a period of a month and yielded quite reproducible results ($\pm 2\%$). Though the PAR 124 is an old model of LIa, its phase drift is small. In normal operation the maximum phase-fluctuation of this model turns out to be less than 0.01°. This phase stability was found particularly valuable to our measurements. Once the phase setting was done at a given frequency, the temperature and voltage scans could be made without any readjustment. It is to be mentioned that the phase resolution of most commercial lock-in amplifiers is around 0.05°. This resolution is not adequate for an accurate phase setting as required in APCS method. However, this will not cause problem in DVCS since the in-phase component was balanced. Moreover, the phase resolution can be improved by a frequency tuning. We had noticed that the phase setting is slightly dependent on frequency. It suggests that instead of using the phase-dial, the fine phase adjustability could be acquired by tuning the reference
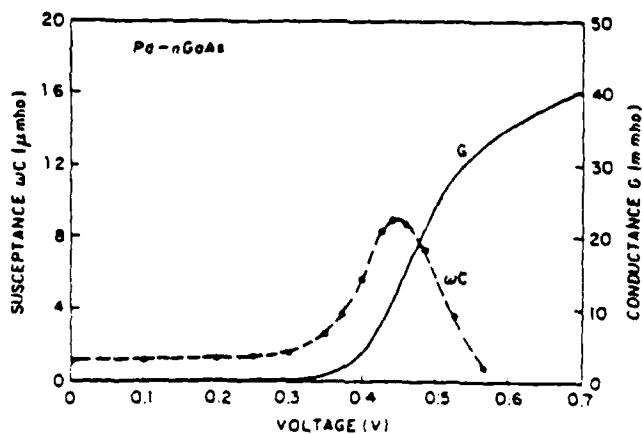
Fig. 3. The susceptance and conductance data obtained from a Pd–nGaAs diode at 108 Hz and 270 K.

frequency. This method was utilized to help set the phase with a resolution beyond that provided by the phase vernier. In order to examine the significance of the phase error, we deliberately turned the phase 1° off its correct position. With the real component carefully nulled out, we found the influence on the output reading to be less than 3%, while the phase accuracy in the experiment could be kept to less than 0.05°.

In Fig. 2, data obtained from the APCS are also presented as a comparison. It can be seen that the results of the DVCS and the APCS are in good agreement at low bias. At high bias the deviation between the two is caused by a phase error. This was confirmed by a phase readjustment in the APCS measurement, such that a 0.005° phase shift could cause the curve to coincide with the DVCS results. Since the DVCS is much less sensitive to the phase error, we believe that this method provides more reliable data. Figure 3 is another example of data obtained from a Pd–n GaAs diode, where the conductance is so high (30 mmho) that overloading becomes a serious problem in APCS. But with DVCS, the capacitance can be measured without

difficulty. It might be worth mentioning that this technique can also be used for $\omega C \gg G$. In that case, one merely puts a variable capacitor in place of the variable resistor, where the capacitive component is to be cancelled. This could be useful in the measurement of conductance in MIS devices.

REFERENCES

1. H. L. Evans, X. Wu, E. S. Yang and P. S. Ho. *Appl. Phys. Lett.* **46**, 486 (1985).
2. C. Barret and A. Vapaille. *Solid-St. Electron.* **18**, 25 (1975).
3. D. W. Greve, *Rev scient. Instrum.* **47**, 1409 (1976).
4. M. Liehr, P. E. Schmid, F. K. LeGoues and P. S. Ho, *Phys. Rev. Lett.* **54**, 2139 (1985).
5. P. S. Ho, E. S. Yang, X. Wu and H. L. Evans. *Phys. Rev. Lett.* **56**, 177 (1986).

The Effect of a Thin UV-Grown Oxide on Metal-GaAs Contacts

M. T. Schmidt, D. V. Podlesnik, H. L. Evans, C. F. Yu, F. S. Yang, and R. M. Osgood, Jr., Microelectronics Sciences Laboratories, Columbia University, New York, NY 10027-6699

Abstract

Photochemistry has been used to grow very thin oxide layers (3 - 10 $\overset{c}{A}$) on GaAs (100) surfaces. Some of these oxides have been annealed to produce Ga-rich oxide surface layers. Electrical characterization of metal contacts deposited on these surfaces show a much greater variation of Schottky barrier heights for the oxidized surfaces, especially for the Ga-rich oxide surfaces. The observed variation is 25% of the GaAs bandgap for Ga-rich oxide surfaces, compared to 8% for clean surfaces. The change in barrier heights is toward the ideal Schottky limit for contacts without interface states. This behavior suggests that the oxidized surface has reduced the density of interface states present in the metal-GaAs contacts.

I. Introduction

Schottky barriers at metal-semiconductor contacts have been the subject of
extensive investigation for several decades. The current understanding indicates
that there are at least two main components which determine a Schottky barrier
height. If an interface contains a high density of electronic states, this will
dominate the barrier formation, allowing differences in the metal-semiconductor
work function to play only a small role in determining the magnitude of a Schottky
barrier [1] This appears to be the case for the GaAs (100) surface where Schottky
barrier heights correlate only weakly with the work function of the metal. If the
GaAs surface can be modified to reduce the density of interface states when a
contact is formed, the Schottky barrier which results would be more influenced by
the metal work function Chemistry at a semiconductor surface is believed to play
an important role in determining the Schottky barrier height of a metal-
semiconductor contact. In this work we have used photochemistry to alter the GaAs
(100) surface through the reaction of oxygen under deep UV illumination prior to
metal deposition and have investigated the electrical characteristics of the
resulting contacts

Previous studies of the effect of thin oxide layers on the electrical properties
of Schottky contacts found that for a variety of metals, the Schottky barrier
increases In those studies the oxide layer was formed by prolonged exposure to
wet oxygen[2] or by an aqueous chemical reaction [3] Oxides formed on GaAs in moist
ambients are different from those grown in dry oxygen probably due to the
incorporation of hydroxyl groups in the oxide [4] Therefore the effect of the
oxides used in our study formed in dry oxygen would be expected to be different
from the previous studies In the previous studies trapped of charge at the

interface is believed responsible for the observed increase in barrier height. Chang, et al.[5] have suggested that a suitably chosen chemical reaction could allow the formation of an oxide on GaAs which would not cause the creation of charge traps.

We have reported that deep UV light greatly enhances the rate of oxidation of GaAs surfaces compared to dark oxidation or oxidation with near UV light exposure.[6] The mechanism responsible for this enhancement is not completely understood, but it is generally agreed that the enhancement is at least in part due to a carrier related effect.[7,8,9] Our recent study indicates that the deep UV enhancement seems to involve the photodissociation of an O_2^- species formed by photogenerated carriers.[10] Thin oxide layers formed in this way have been found to be stoichiometric with respect to Ga and As. Oxide layers which are nonstoichiometric have also been formed by annealing after oxide growth. This results in a surface which is Ga-oxide rich, as will be shown later. Metal contacts were deposited on both stoichiometric oxide and Ga-oxide rich surfaces. The variation in Schottky barrier height for various metal contacts formed on oxidized surfaces was found to depend much more strongly on the work function of the metal overlayer than is observed on clean GaAs surfaces. It is concluded that a more ideal relationship between metal work function and Schottky barrier height exists for the oxidized surface than occurs for clean GaAs surfaces.

II Experiment

The experiments in this study were performed in a two chamber UHV surface analysis system which allowed UV light exposure of the sample and in situ metal deposition. Metal deposition was from an electron beam evaporation source where

4

the source to sample distance is 30 cm. The GaAs used in this study was (100) oriented, nominally undoped with $n = 1 \times 10^{16}$ cm^{-3}. All samples were degreased by sequentially soaking in hot trichlorethylene, acetone, and methanol. They were then chemically etched for 15 seconds in $NH_4OH:H_2O_2:H_2O$ (1:1:10), oxidized in $H_2SO_4:H_2O_2:H_2O$ (10:1:1) for 1 minute, quenched in $10^{20}\Omega$-cm water, and blown dry with N_2. A sample was then mounted on a Mo plate with melted In and immediately inserted into the UHV system, where it was heated in vacuum to 550 °C for 5 minutes. This desorbed the oxide layer formed during cleaning and served to anneal the In on the back of the sample, providing an ohmic contact for electrical measurements. Waldrop has performed LEED studies to show that a similar cleaning procedure provides a clean, well ordered surface.[11]

After the surface had been cleaned and XPS performed, very pure O_2 was leaked into the preparation chamber, typically to a pressure of 10 Torr. Details of the surface oxidation procedure have been previously reported.[4,6] Deep UV irradiation of the GaAs was performed with a low power, high repetition rate KrF excimer laser (λ = 248 nm) of photon energy 4.9 eV. Typical laser power used is < 2 mJ/cm^2 so that thermal effects are negligible. The GaAs was irradiated at 100 Hz for 5 min to 20 min to produce the desired oxide thickness.

Oxide thicknesses were measured using XPS by determining the ratio of areas of the oxidized As peak to the As peak from the substrate.[6] Oxide thicknesses between 3 and 50 Å can be accurately controlled by varying the O_2 pressure and duration of irradiation. A 3 Å oxide can be reproducibly formed with a 5 min irradiation, while a 15 Å oxide will form in 20 min of irradiation.

The oxides used in this study are between 3 and 10 Å thick in order to produce contacts with good ideality factors. The oxides grown in the manner described above are found by XPS to be stoichiometric with respect to Ga and As, consisting

predominantly of Ga_2O_3 and As_2O_3.[6] However, we have also formed nonstoichiometric oxides. Using the known fact that Ga_2O_3 is more thermally stable than As_2O_3, we can produce a Ga-oxide rich surface by annealing the oxidized GaAs. After a stoichiometric oxide is grown, and the chamber has been re-evacuated, the sample is heated to 500°C[12] for 10 min and allowed to slowly cool. This is believed to cause a reaction where As_2O_3 reacts with the substrate to form Ga_2O_3 while As desorbs from the surface.

Typical XPS data is shown in Fig 1 comparing a surface with a stoichiometric oxide to the same surface after annealing The As $2p_{3/2}$ peak indicates that the oxide layer is 6 A thick before annealing. Notice that after annealing there is very little signal from the chemically shifted peak due to oxidized As Since the Ga peak cannot be used to determine the thickness of the oxide layer, we used the preannealed oxide thickness as a measure of the amount of oxygen on the GaAs surface before metal deposition. The Ga peak after annealing, however, is still chemically shifted 0.9 eV from the substrate peak position, characteristic of oxidized Ga. The ratio of the As $2p_{3/2}$ signal to the Ga $2p_{3/2}$ signal for the annealed surface is 0.46, in agreement with the prediction that the surface is Ga rich. The ratio for a stoichiometric surface has been found from MBE prepared samples to be 1.4 [13] Notice in Fig 1 that the O 1s peak is unchanged in magnitude before and after annealing, consistent with our assumption that the amount of oxygen on the surface does not change during annealing

The metals used to form contacts in this study were Pd, Cu, Cr, and Ti They were chosen because they have different reactivities with GaAs oxides, they cover a range of metal work function values, and all have been found to react strongly with GaAs surfaces [14,15,16] Pd and Cu are inert to an oxide layer on GaAs, while Ti and Cr react completely with the oxygen in thin oxide layers [17] The contacts were

formed by evaporation through a thin Mo mask pulled close to the sample surface just prior to deposition. The contacts were dots of 8×10^{-3} cm^2 area with at least six contacts per sample. After metal deposition, the sample was removed from the UHV system and the contacts were electrically characterized by dark current-voltage (I-V) and internal photoemission (PE) measurements.[18]

III. Results

The results of electrical measurements for the various metals on the different surface oxides, as well as on clean control samples, are summarized in table 1. The results are typical values from measurements of several contacts of each category. The optimum oxide thickness to cause the maximum change in Schottky barrier height for each metal was not investigated. It can be seen that an oxide layer reduced the barrier height for Cu, Cr, and Ti, but increased the barrier for Pd. It will be discussed below that this correlates with the work functions of the various metals. It is also clear that the contacts with Ga-rich oxide layers produced a larger barrier height variation for the various metals and better ideality factors from I-V measurements. The better ideality factor is especially noticeable for the Cu contacts. For Cu contacts, similar stoichiometric oxide thickness and Ga-rich oxide thickness result in a much better ideality factor for the Ga-rich oxide contact.

The barrier heights in table 1 measured by internal photoemission agree with the trends observed in the I-V measurements. This is important since PE measurements are the most direct measure of barrier heights[18] not dependant on contact area, physical constants, or low barrier height regions. In some cases PE measurements could not be made. The variation in PE barrier heights on clean surfaces for

different metals is only 0.15 eV, while on Ga-rich oxide surfaces the variation is 0.28 eV. The increase in barrier height variation is even greater from I-V measurements. From these it is seen to vary 0.39 eV for Ga-rich oxide surfaces, compared to only 0.12 eV for contacts on clean surfaces. This corresponds to 25% of the GaAs bandgap, compared to 8% for the clean surfaces examined here.

I-V data from each of the metals and surface treatments is shown in Figs. 2 and 3. For the metals reactive with oxygen, Cr and Ti, the shift of the I-V plot is very similar, see Fig 2. The stoichiometric oxide shifts the current significantly higher, while the shift for a Ga-rich oxide is even greater These shifts indicate a decrease in Schottky barrier height. In all cases, the plots are parallel, indicating no change in contact ideality. It was found that the oxide thickness had less of an effect on the contact ideality factor for metals reactive with oxygen than for metals inert to the oxide layer. This could be of practical importance in making contacts to oxidized surfaces. For the Cu contact, Fig 3(a), a stoichiometric oxide only slightly shifts the plot to higher current, indicating lower barrier, but it also degrades the contact ideality Cu contact idealities were the most strongly affected by the stoichiometric oxide layers of the contacts in this study. For a Ga-rich oxide, however, the plot shifts to that for a much lower barrier with good ideality.

The behavior for Pd contacts is the opposite of that for the above metals. Fig 3(b) shows that a stoichiometric oxide shifts the I-V plot downward corresponding to a higher barrier height. This plot indicates no degradation of contact ideality, but this is true only for oxides less than 7 A thick. Thicker oxides increase the ideality factor for the contact, and produce lower apparent barrier heights. Fig 3(b) also shows a further increase in barrier height for Pd on a Ga

rich oxide surface with the ideality factor restored to the original value. While this further increase in barrier height is slight, it was consistently observed.

IV. Discussion

The large increase in the range of Schottky barrier heights for contacts formed on surfaces which have been oxidized using deep UV irradiation at room temperature suggests that the density of interface states has been reduced, allowing the metal work function to play a larger role in the barrier formation. In fig. 4 the Schottky barrier height for the various contacts have been plotted versus the work function of the contact metal. Also shown are the barrier height expected if the Fermi level is completely pinned and the ideal barrier height limit in the absence of interface states. In all cases, the change in barrier height for contacts on oxidized surfaces is toward the ideal limit, that is, in the direction suggesting a decrease in interface state density. Again, this change is most extreme for the Ga-rich oxide surfaces, indicating a larger decrease in interface state density for this case. This is also supported by the better ideality factors noted above for contacts on Ga-rich oxide surfaces, since a high density of interface states will result in a higher ideality factor from current-voltage measurements.

A GaAs surface with a Ga-rich oxide has been previously found by XPS to have less surface band bending than a surface with a stoichiometric oxide.[19] The reduced band bending at the surface was attributed to a reduction in the interface state density. It was suggested that this might occur when the stoichiometric oxide was annealed and Ga_2O_3 was formed in the reaction with the substrate. In our data, the interface state density appears to be reduced even for the stoichiometric oxide, but reduced further for the Ga-rich oxide. This may be due to the

nonbonding, occupied orbital associated with the As_2O_3 molecule.[20]   This bonding defect can act as a filled acceptor state on the surface. The reduction of the amount of As_2O_3 on the surface when a Ga-rich oxide is formed may be the reason for the larger variation of Schottky barrier heights observed for contacts to annealed oxides.

Our results are also in agreement with the recent unpinning of the GaAs surface using a photochemical process in flowing water. A possible explanation for the unpinning was a passivating Ga-oxide layer on the surface.[21] Our study, however, is the first to form Schottky contacts on a well characterized Ga-oxide rich surface where an increased variation in Schottky barrier height is demonstrated.

The similarity of behavior for metals reactive with the surface oxide layer, and metals inert to the surface oxide layer, indicates that chemical reactions involving the metal and the oxygen at the surface are not the dominant cause for the apparent reduction in interface states. It is not clear, however, whether the increased Schottky barrier variation observed here is due to a reduction of surface states before metal deposition, or a reduction in the interface states formed during metal deposition Oxygen at the surface may play a role in inhibiting alloy formation between Ga, As, and the metal overlayer.[22] Oxide compounds also give the semiconductor surface a more ionic character which would tend to decrease the density of the induced states when a metal is deposited.[23,24] This is consistent with the observed increased reduction in interface states for a Ga-oxide rich interface since Ga_2O_3 has a more ionic character than As_2O_3.

V. Conclusion

A much wider variation in Schottky barrier heights for contacts on GaAs (100) surfaces which had been oxidized using low intensity, deep UV illumination at room temperature was observed compared to contacts on clean GaAs surfaces. This has been interpreted as resulting from a reduction of interface states in the metal-GaAs contact, since the Schottky barrier which results for metals deposited on the surfaces with oxides varies more strongly with the work function of the metal than Schottky barriers on clean GaAs surfaces. This reduction in interface states is especially pronounced for a stoichiometric oxide which has been annealed to form a Ga-rich oxide surface.

Acknowledgement

Figure Captions

Figure 1   X-ray Photoelectron Spectroscopy data for GaAs surfaces (a) with stoichiometric oxides, and (b) the same surface after annealing 10 min at 500°C.   The oxidized As signal is shifted to lower kinetic energy. The curves labeled (b) show a Ga oxide rich surface.   Note the O 1s peak has the same magnitude before and after annealing.   The data has been smoothed and the background has been subtracted in all cases.

Figure 2   Natural log of the current versus applied voltage for the metals reactive with a GaAs-oxide layer.   (a) Ti on n-GaAs with the indicated surface treatments.   (b) Cr on n-GaAs with the indicated surface treatments

Figure 3   Natural log of the current versus applied voltage for the metals inert to a GaAs-oxide layer.   (a) Cu on n-GaAs with the indicated surface treatments.   (b) Pd on n-GaAs with the indicated surface treatments.

Figure 4   Schottky Barrier Height versus metal work function   Also plotted is the ideal Schottky limit for a barrier in the absence of interface states (given by the difference between the metal work function and the electron affinity of GaAs) and a probable pinned limit if barrier formation is dominated by interface states   Points labeled Cl are contacts to clean surfaces. Points labeled ST are contacts to UV grown, stoichiometric oxides.   Points labeled GR are contacts to Ga-rich oxide surfaces.   (Metal work functions from H.B. Michaelson, J Appl Phys. 48.4729 (1977). Figure after C.R. Crowell, H.B. Shore, and E.F. LaBate. J. Appl. Phys. 36. 3843 (1969))

Table 1.

Metal	Surface Preparation	Oxide Thickness (A)	PE Barrier (eV)	I-V Barrier (eV)	Ideality factor
Pd	Ref.			.87	1.03
	Clean	0	.85	.86	1.03
	Stoichiometric Oxide	3.0	-	.92	1.04
	Stoichiometric Oxide	5.0	.87	.92	1.04
	Ga-rich Oxide	6.8	.93	.93	1.04
Cu	Ref.			.92	1.05
	Clean	C	-	.88	1.05
	Stoichiometric Oxide	4.2	.84	.84	1.24
	Ga-rich Oxide	4.2	.65	.77	1.03
Ti	Ref.			.79	1.03
	Clean	0	.74	.78	1.02
	Stoichiometric Oxide	7.2	.72	.65	1.02
	Ga-rich Oxide	7.1	.70	.63	1.00
Cr	Ref.			.73	1.04
	Clean	0	.70	.73	1.05
	Stoichiometric Oxide	7.5	.68	.62	1.08
	Ga-rich Oxide	6.0	-	.54	1.07

Ref. values from J.R. Waldrop, Appl.Phys.Lett. 44,1002(1984); J.R. Waldrop, J. Vac. Sci. Technol. B2, 445 (1984).

Oxide thickness $^{+}-$ 10% ; I-V and PE barriers $^{+}-$ .01 eV.

References

1.   See, for example, Heinz K. Henisch, Semiconductor Contacts. (Oxford University Press, New York, 1984).

2.   R. B. Childs, J. M. Ruths, T. E. Sullivan, and S. J. Fonash. J. Vac. Sci. Technol. 15, 1397 (1978).

3.   R. L. Van Meirhaeghe, W. H. Laflere, and F. Cardon. Solid-State Electron. 25, 1089 (1982).

4.   C.F. Yu, M.T. Schmidt, D.V. Podlesnik, and R.M Osgood,Jr., J. Vac. Sci. Technol B5, 1087 (1987).

5   R. P. H Chang J. J. Coleman, A. J. Polak, I. C Feldman and C. C. Chang Appl. Phys. Lett. 34, 237 (1979).

6   C.F. Yu, D.V. Podlesnik, M.T. Schmidt, H.H. Gilgen, and R M Osgood, Jr., Chem. Phys. Lett. 130, 301 (1986)

7.   S.A. Schafer and S.A Lyon, J. Vac. Sci Technol 19, 494 (1981)

8.   W.G. Petro, I Hino, S. Eglash, I. Lindau, C.Y. Su, and W.F Spicer, J. Vac Sci Technol. 21, 405 (1982)

9.   F. Bartels and W. Mnch, Surf Sci. 143, 315, (1984)

10.   C. F. Yu, M.T. Schmidt, D. V. Podlesnik, F. S. Yang and R M Osgood, Jr., These proceedings

11.   J.R. Waldrop, Appl Phys Lett 44, 1002 (1984), J.R Waldrop J Vac Sci. Technol B2, 445 (1984)

12.   This is the nominal temperature on our sample rod controller The actual sample temperature could be up to 10%. lower.

13.   J.M. Woodall, P. Oelhafen, T.N. Jackson, J.I. Freeouf and G.D Pettit, J. Vac Sci Technol. B1, 795 (1983).

14   R. Ludeke and G. Landgren, Phys. Rev. B 33, 5526 (1986)

15   S.H Pan, T. Kendelewicz, W.G Petro, M.D. Williams I Lindau and W.F. Spicer, Material Research Society 1983 Symposium. (Elsevier, New York, 1984), Vol 25, p 335

16   M.D Williams, T. Kendelewicz, R.S List, N. Newman, C I McCants, I. Lindau, and W.F Spicer, J. Vac Sci Technol B3 122 (1985)

17. Steven P. Kowalczyk, J.R. Waldrop, and R.W. Grant, Appl. Phys. Lett. **38**, 167 (1981); Steven P. Kowalczyk, J.R. Waldrop, and R. W. Grant, J. Vac. Sci. Technol. **19**, 611 (1981); M.T. Schmidt, C.F. Yu, F.S. Yang, and R.M. Osgood,Jr., unpublished results.

18. See, for example, S.M. Sze, <u>Physics of Semiconductor Devices</u> (Wiley, New York, 1981).

19. R.W. Grant, S.P. Kowalczyk, J.R. Waldrop, and W.A. Hill, in <u>Physics of MOS Insulators</u>, edited by G. Lucovsky, S.T. Pantelides, and F.L. Galeener (Pergamon Press, New York, 1908) p. 202.

20. G. Lucovsky and R.S. Bauer, J. Vac. Sci. Technol. **17**, 946 (1980).

21. S.D. Offsey, J.M Woodall, A.C. Warren, P.D. Kirchner, T.I Chappell, and G.D. Pettit, Appl Phys. Lett. **48**, 475 (1986)

22 J.S Solomon and S.R. Smith, J. Vac. Sci. Technol A**5**, 1809 (1987)

23 S. Kurtin, T.C. McGill, and C. A. Mead. Phys Rev Lett **22**, 1433 (1969).

24. L.J. Brillson, J. Vac. Sci. Technol. **15**, 1378 (1978).

As2p$^2/_3$ Ga2p$^2/_3$

O 1s

XPS INTENSITY (ARBITRARY UNITS)

(a)

(b)

(a)

(b)

156   168 362   374
KINETIC ENERGY (eV)

(a)

(b)

945   965
KINETIC ENERGY (eV)

CURRENT (A)

10^{-10} 10^{-8} 10^{-6} 10^{-4} 10^{-2}

Ti/n-GaAs

.120 .240 .360 .480 .600

VOLTS

CLEAN SURFACE
STOICHIOMETRIC OXIDE
Ga-RICH OXIDE

CURRENT (A)

10^{-10} 10^{-8} 10^{-6} 10^{-4} 10^{-2}

Cr/n-GaAs

.120 .240 .360 .480 .600

VOLTS

CLEAN SURFACE
STOICHIOMETRIC OXIDE
Ga-RICH OXIDE

CURRENT (A)

Cu/n-GaAs

VOLTS

CLEAN SURFACE
STOICHIOMETRIC OXIDE
Ga-RICH OXIDE

CURRENT (A)

Pd/n-GaAs

VOLTS

CLEAN SURFACE
STOICHIOMETRIC OXIDE
Ga-RICH OXIDE

SCHOTTKY BARRIER – eV

METAL WORK FUNCTION – eV

IDEAL
SCHOTTKY
LIMIT
$(\phi_m - X_{GaAs})$

CONVERSION OF POISSON PHOTONS INTO SUB-POISSON PHOTONS BY THE ACTION OF ELECTRON FEEDBACK

Federico Capasso and Malvin C. Teich

Conversion of Poisson Photons into Sub-Poisson Photons
by the Action of Electron Feedback

Federico Capasso

AT&T Bell Laboratories, Murray Hill, New Jersey 07974

and

Malvin C. Teich

Columbia Radiation Laboratory, Department of Electrical Engineering, Columbia University, New York, New York 10027

(Received 12 May 1986)

Poisson photons may be converted into sub-Poisson (amplitude-squeezed) photons by the action
of an electron current configured in an external feedback loop. The generation mechanism in-
volves single-photon transitions so that the source can be made arbitrarily sub-Poissonian. Non-
linear optics is not invoked. A useful configuration involves a photon emitter illuminating a
detector-source combination in a closed-loop system. Two solid-state implementations of the
detector-source combination are suggested.

It is by now well known that nonclassical light can be
generated in the laboratory. Such light may exhibit a
variety of exotic properties including sub-Poisson pho-
ton statistics,[1,2] antibunching[3] (sub-Poisson photon
statistics and antibunching need not necessarily accom-
pany each other[4]), and squeezing.[5] Sub-Poisson light
is also called amplitude-squeezed light.

The use of an external feedback system[6] for the gen-
eration of sub-Poisson light was first suggested by ex-
periments in which feedback was used to produce
sub-Poisson *electrons*.[7-9] Both the experiment of
Walker and Jakeman[8] and the experiment of Machida
and Yamamoto[9] involved laser (Poisson) photons il-
luminating a photodetector and an electronic negative
feedback path from the detector to the source. In the
former experiment the feedback directly controlled the
photons at the output of the laser, whereas in the latter
experiment the feedback controlled the current at the
input to the laser. Nevertheless, the principle in-
volved in the two experiments is the same.[6] Unfor-
tunately, these simple configurations could not gen-
erate usable sub-Poisson *photons* since the feedback
current is generated from the annihilation of the in-
loop photons. However, under special circumstances
an external feedback system can be used to produce
sub-Poisson photons, such as when correlated photon
pairs are available[10-12] or a quantum nondemolition
measurement may be made.[13] The resultant light may
only be weakly sub-Poissonian in such cases because
these multiphoton processes involve a series of detec-
tions and/or a weak nonlinear effect.

In this Letter we propose a new approach for the
conversion of Poisson photons into sub-Poisson pho-
tons *via single-photon transitions*. There does not ap-
pear to be any fundamental limit that would impede

the technique from being used to produce an arbitrari-
ly intense cw light source that is also arbitrarily sub-
Poissonian. It makes use of the action of an electron
current configured in a feedback loop. Consider, for
example, an optical system in which a photon emitter
illuminates a detector-source combination, in a
closed-loop circuit. Two alternative configurations are
shown in Fig. 1. The character of the photon emitter
is immaterial; we have chosen it to be a light-emitting
diode (LED) for simplicity, but it could be a laser.[6] In
Fig. 1(a) the photocurrent derived from the detection
of light from the LED photon emitter is negatively fed
back to the LED input. It has been established both
theoretically[6] and experimentally[9] that, in the absence
of the block labeled "source," sub-Poisson electrons
will flow in a circuit such as this. This conclusion is
also generally valid in the presence of this block,
which simply acts as an added impedance to the elec-
tron flow. Incorporating this element into the system
critically alters its character, however, since it permits
the sub-Poisson electrons flowing in the circuit to be
converted into sub-Poisson photons by means of elec-
tron transitions. The key to the achievement of this
effect is the replacement of the detector used in other
feedback configurations with a structure that acts
simultaneously as a detector and a source. The elec-
trons simply emit sub-Poisson photons and continue
on their way. In the absence of the feedback path, of
course, the electrons would simply emit Poisson pho-
tons. Thus, the introduction of the electron feedback
converts Poisson photons into sub-Poisson photons.
The configuration in Fig. 1(b) is similar except that
the (negative) feedback current modulates (gates) the
light intensity at the output of the LED rather than the
current at its input.

(a)

(b)
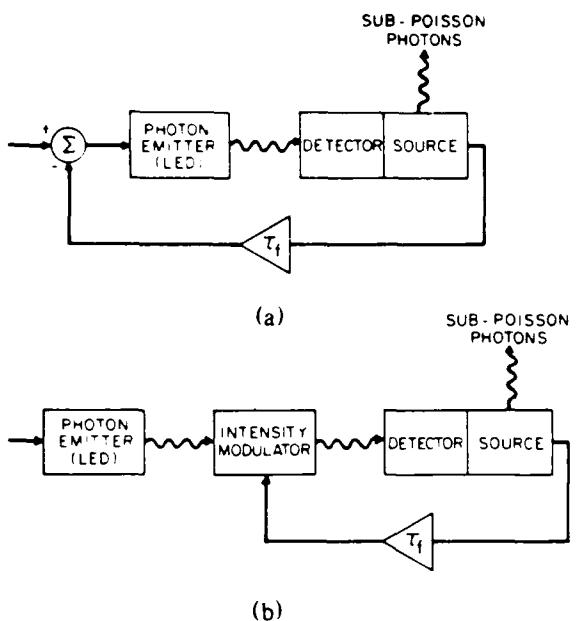
FIG. 1. Generation of sub-Poisson photons by means of negative feedback. The feedback produces sub-Poisson electrons in the detector-source which, in turn, generate sub-Poisson photons. (a) Negative feedback modulating the photon-emitter input current. (b) Negative feedback modulating the photon-emitter output light. Wavy lines represent photons; solid lines represent electron current. τ_f signifies the feedback time constant.

In Fig. 2 we illustrate two possible solid-state detector-source configurations. The basic structure consists of a reverse-biased p^+-i-n^+ diode where the p^+ and n^+ heavily doped regions have wider band gaps than the high-field, light-absorbing and -emitting i region. This arrangement ensures both high quantum efficiency at the incident photon wavelength (to which the p^+ window layer is transparent) and high collection efficiency (due to the waveguide geometry) for the light generated by the electrons drifting in the i layer. An edge-emitting geometry is therefore appropriate. To maximize the collection efficiency, some of the facets of the device could be reflectively coated. Two light-generation schemes are explicitly considered here: single-photon dipole electronic transitions between the energy levels of the quantum wells [Fig. 2(a)] and impact excitation of electroluminescent centers in the i region by drifting electrons [Fig. 2(b)].

In the first scheme [Fig. 2(a)], the i region consists of a layer in which the incident photons are absorbed (detection region) and an adjacent quantum-well region in which photons are generated by sequential resonant tunneling (source region). Capasso, Mohammed, and Cho[14] have recently demonstrated the sequential resonant tunneling of electrons through a

(a)

(b)
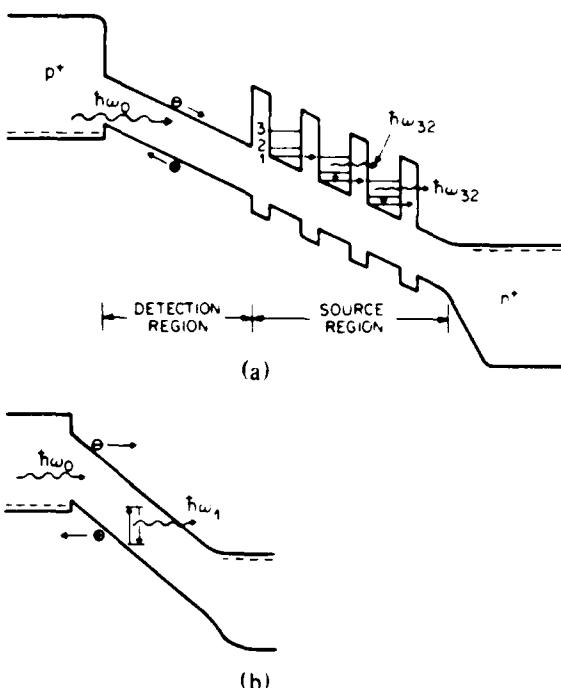
FIG. 2. (a) Band diagram of the quantum-well detector-source. The energy of the incident photon emitted by the LED is denoted $\hbar\omega_0$. The absorbing region (detection region) is of $Ga_{0.47}In_{0.53}As$ or GaAs, typically 1 μm thick. In the source region, the wells are $Ga_{0.47}In_{0.53}As$ or GaAs in the thickness range 150–300 Å. The barrier layers are of $Al_{0.48}In_{0.52}As$ (in the case of $Ga_{0.47}In_{0.53}As$ wells) or AlAs (in the case of GaAs wells) and should be in the thickness range 20–50 Å to achieve tunneling times < 1 psec. Photons of energy $\hbar\omega_{32}$ are emitted via transitions from level 3 to 2. The p^+ and n^+ wide-gap regions are of $Al_{0.48}In_{0.52}As$ (or AlAs). The interface between the i and n^+ regions is compositionally graded. (b) Energy-band diagram of a detector-source with electroluminescent centers that are impact excited by energetic photoelectrons, emitting photons with energy $\hbar\omega_1$.

quantum-well superlattice consisting of 35 periods of $Al_{0.48}In_{0.52}As$ (140 Å) and $Ga_{0.47}In_{0.53}As$ (140 Å). The observation of this effect, first predicted in 1971 by Kazarinov and Suris,[15] has been made possible by the remarkable quality of superlattice structures recently achieved by the molecular-beam-epitaxy growth technique.

The structure presented in Fig. 2(a) is similar to that previously used in the resonant tunneling experiment.[14] The reverse-bias voltage applied to the detector is adjusted such that the potential-energy drop across the superlattice period (barrier plus well) is equal to the energy difference between the bottom of the third and first subbands of the quantum wells. Such discrete subbands (which have momentum

dispersion in the plane of the layer) arise from size quantization perpendicular to the layers when the electron de Broglie wavelength is comparable to the well thickness.[16] Under such bias conditions electrons undergo resonant tunneling into the third energy level of the quantum wells and the device current exhibits a peak due to the enhanced tunneling probability. Once the electron finds itself in the third energy level, it relaxes by the emission of either phonons or infrared photons.

The relative probability of these two processes is an important parameter of the system because a nonradiative decay represents random deletion and this reduces the sub-Poisson character of the generated light.[17] The most important nonradiative process is scattering, by intersubband optical phonons, of the electron located at the bottom of the third energy level. This results in the electron finding itself in the $n = 2$ subband after the emission or absorption of a phonon. (Once the electron has scattered to the second subband it nonradiatively relaxes to the bottom of the ground-state subband.) The intersubband optical-phonon–assisted electron transitions are much stronger than the acoustic-phonon–assisted transitions and have a rate which is typically several orders of magnitude greater than the radiative spontaneous emission rate from the $n = 3$ to the $n = 2$ level (this is estimated to be $\approx 10^{7}-10^{8}$ sec^{-1} at a wavelength $\lambda = 10$ μm, in the 140-Å well structures considered in Ref. 14). This radiative transition is dipole in nature and the emitted light is polarized normal to the plane of the layers. These photons cannot be reabsorbed by ground-state electrons since $E_3 - E_2 \neq E_2 - E_1$. Indeed, this dipole transition has recently been observed in absorption between the ground and first-excited states of quantum wells.[18]

The probability that an electron undergoes a radiative transition (i.e., the radiative efficiency) η_r at the emission frequency $(E_3 - E_2)/h$ can nevertheless be increased substantially by quenching of the optical-phonon–assisted transition. This may be accomplished by an appropriate increase of the well thickness thereby making the energy separation $E_3 - E_2$ smaller than the optical-phonon energy (≈ 35 meV), in which case the device would emit in the middle infrared wavelength region. Another method to increase the radiative efficiency is to apply a magnetic field perpendicular to the layers. In the presence of the magnetic field, the electron states in the plane of the layer become quantized resulting in a manifold of discrete Landau levels. This total quantization of the electron wave function dramatically reduces the available density of states for scattering, thus quenching the intersubband transitions. Recent experiments by Ryan et al.[19] have indeed shown that in the presence of a perpendicular magnetic field, the energy relaxation rate of an electron-hole plasma in quantum wells is reduced by many orders of magnitude. Although it is difficult to estimate precisely the radiative efficiency obtainable with the selective quenching of optical-phonon–assisted intersubband transitions and/or with the application of an external magnetic field, we expect that values between 0.01 and 0.1 should be achievable. We have applied the standard four-level–system rate equations to estimate that 1-10 nW of infrared power at a wavelength ≥ 10 μm should be achievable for a photocurrent $\approx 10^{-5}$ A. Another suggested configuration, making use of electroluminescent centers impact excited by energetic photoelectrons,[20] is presented in Fig. 2(b). The use of a superlattice configuration in the i region could be useful in the enhancement of the impact-excitation probability.[21]

The ability of configurations such as these to generate sub-Poisson light requires a number of interrelations among various characteristic times associated with the system. Specifically, the condition $\tau_f \gg \tau_{tr}, \tau_{resp}$ must be obeyed, where τ_f is the feedback time constant, τ_{tr} is the transit time of the electrons and holes through the structure, and τ_{resp} is the response time of the detector-source. This condition ensures that the detector-source response is fast enough to be properly integrated by the feedback circuit. It is also required that $T \gg \tau_f$, where T is the counting (integration) time of the external detector that monitors the sub-Poisson light. This ensures that the photons are monitored for a time that is sufficiently long for the negative feedback to act on the rate of LED photon emissions. Finally, it is also required that $T \gg \tau_r$, where τ_r is the spontaneous-emission lifetime of the radiative transition. This assures that the photon emissions will be captured in the proper counting time interval.[2,22]

An estimate of the degree to which a light source is sub-Poisson (amplitude squeezed) is provided by the ratio of the photon-number variance to the photon-number mean (Fano factor), $F_n(T) = \text{Var}(n)/\langle n \rangle$. For sub-Poisson light, the condition $0 \leq F_n(T) < 1$ is obeyed; the closer $F_n(T)$ is to zero, the more sub-Poissonian is the light. The mechanism for the generation of sub-Poisson light described in this Letter can be characterized by a sub-Poisson electron counting process e, each event of which independently generates a random number of photons M in the source. The overall photon-number Fano factor $F_n(T)$ can then be represented in terms of the Fano factor for the electron number $F_e(T)$ and the Fano factor for the source random variable $F_M(T)$. The relationship is[17,22]

$$F_n = \langle M \rangle F_e + F_M. \tag{1}$$

where $\langle M \rangle$ is the average number of photons generat-

ed in the source by each electron.

For the case at hand it is reasonable to assume that the source random variable is Bernoulli distributed in each stage of the device,[22] with the probability of an electron giving rise to a photon denoted η_r. No generality is lost by consideration of the multilayer superlattice case, which consists of m independent stages. The source statistics will then take the form of a binomial random variable with $\langle M \rangle = m\eta_r$ and $\mathrm{Var}(M) = m\eta_r(1 - \eta_r)$. In the presence of random deletion arising from other factors (e.g., finite geometrical photon-collection efficiency, absorption, external detection) and background or dark photons, these results remain valid upon the replacement of η_r by the quantity $\eta\beta$, where η is the overall quantum efficiency from electrons to detected photons and β is a factor representing the admixture of independent dark and/or background events.[17] In that case, F_n will be the Fano factor for the *detected* photons. (It may be useful to operate structures such as those discussed here at reduced temperatures to assure that $\beta \approx 1$ and to enhance the resonant tunneling current.) Equation (1) then provides

$$F_n - 1 = \eta\beta[mF_e - 1]. \tag{2}$$

From Eq. (2) it is evident that sub-Poisson behavior may be discerned when $F_e < 1/m$. However, the lowest Fano factor is achieved when $m = 1$. In this case, the photon counting process is simply a randomly deleted version of the electron counting process so that Eq. (2) reduces to its usual familiar form.[2,17,22]

Numerical estimates can be obtained for the degree of sub-Poisson behavior that is expected to be observable for the two structures discussed here, on the assumption that $\beta = 1$. For the superlattice device, η will be the product of the quantum efficiency of the *external* detector ($\eta_d \approx 0.8$), the geometrical collection efficiency of the emitted sub-Poisson photons ($\eta_g \approx 0.5$), and the radiative efficiency ($\eta_r \approx 0.1$). From the experiment of Machida and Yamamoto[9] we estimate that $F_e \approx 0.2$ (which is principally limited by the quantum efficiency of the detector in the detector-source combination). Thus, for this particular superlattice structure, Eq. (2) provides an overall Fano factor $F_n \approx 0.968$ for the detected sub-Poisson photons. The estimated Fano factor for the electroluminescent structure falls in the same range. These estimates provide a significant potential improvement over the value observed in the space-charge–limited Franck-Hertz experiment.[2] Lower values of the Fano factor can be achieved in structures that exhibit higher radia-

tive efficiency. As indicated earlier, there is no fundamental limit that impedes this scheme from being used to produce an arbitrarily sub-Poissonian cw light source of arbitrarily high intensity.

[1]R. Short and L. Mandel, Phys. Rev. Lett. 51, 384 (1983).

[2]M. C. Teich and B. E. A. Saleh, J. Opt. Soc. Am. B 2, 275 (1985).

[3]H. J. Kimble, M. Dagenais, and L. Mandel, Phys. Rev. Lett. 39, 691 (1977).

[4]M. C. Teich, B. E. A. Saleh, and D. Stoler, Opt. Commun. 46, 244 (1983).

[5]R. E. Slusher, L. W. Hollberg, B. Yurke, J. C. Mertz, and J. F. Valley, Phys. Rev. Lett. 55, 2409 (1985).

[6]J. H. Shapiro, M. C. Teich, B. E. A. Saleh, P. Kumar, and G. Saplakoglu, Phys. Rev. Lett. 56, 1136 (1986).

[7]M. C. Teich and G. Vannucci, J. Opt. Soc. Am. 68, 1338 (1978).

[8]J. G. Walker and E. Jakeman, Proc. Soc. Photo-Opt. Instrum. Eng. 492, 274 (1985).

[9]S. Machida and Y. Yamamoto, Opt. Commun. 57, 290 (1986).

[10]B. E. A. Saleh and M. C. Teich, Opt. Commun. 52, 429 (1985).

[11]E. Jakeman and J. G. Walker, Opt. Commun. 55, 219 (1985).

[12]C. K. Hong and L. Mandel, Phys. Rev. Lett. 56, 58 (1986).

[13]Y. Yamamoto, N. Imoto, and S. Machida, Phys. Rev. A 33, 3243 (1986).

[14]F. Capasso, K. Mohammed, and A. Y. Cho, Appl. Phys. Lett. 48, 478 (1986).

[15]R. F. Kazarinov and R. A. Suris, Fiz. Tekh. Poluprovodn. 5, 797 (1971), and 6, 148 (1972) [Sov. Phys. Semicond. 5, 707 (1971), and 6, 120 (1972)].

[16]T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. 54, 437 (1982).

[17]M. C. Teich and B. E. A. Saleh, Opt. Lett. 7, 365 (1982).

[18]L. C. West and S. J. Eglash, Appl. Phys. Lett. 46, 1156 (1985).

[19]J. I. Ryan, R. A. Taylor, A. J. Turberfield, and J. M. Worlock, Physica (Amsterdam) 134B, 318 (1985).

[20]J. Shah and A. F. DiGiovanni, Appl. Phys. Lett. 33, 995 (1978).

[21]C. J. Summers and K. F. Brennan, Appl. Phys. Lett. 48, 806 (1986).

[22]M. C. Teich, B. E. A. Saleh, and J. Peřina, J. Opt. Soc. Am. B 1, 366 (1984).

EXCESS NOISE FACTORS FOR CONVENTIONAL AND SUPERLATTICE
AVALANCHE PHOTODIODES AND PHOTOMULTIPLIER TUBES

MALVIN C. TEICH, KUNIAKI MATSUO, and BAHAA E. A. SALEH

Excess Noise Factors for Conventional and Superlattice Avalanche Photodiodes and Photomultiplier Tubes

MALVIN C. TEICH, SENIOR MEMBER, IEEE, KUNIAKI MATSUO, MEMBER, IEEE, AND
BAHAA E. A. SALEH, MEMBER, IEEE

Abstract—Light falling on a photodetector produces an output current that fluctuates. The noise in this signal arises from two sources: randomness in the photon arrivals and randomness in the carrier multiplication process intrinsic to the photodetector. A general formula is derived for the variance of the photodetector output current in terms of parameters characterizing these two sources of randomness (the photon-number variance-to-mean ratio for the light and the excess noise factor for the detector). An important special case of this formula illustrates that the output-current variance is directly proportional to the detector excess noise factor when the number of photons at the input to the detector is Poisson distributed. Explicit expressions for excess noise factors are provided for three kinds of photodetectors: the double-carrier conventional avalanche photodiode, the double-carrier superlattice avalanche photodiode, and the photomultiplier tube. The results for the double-carrier superlattice device are new; it is shown that even a small amount of residual hole ionization can lead to a large excess noise factor. Comparisons are drawn among the detectors in terms of their noise properties.

I. INTRODUCTION

LIGHT falling on a photodetector produces an output current that fluctuates. The noise in this signal arises from two sources: randomness in the photon arrival number and randomness in the carrier multiplication process intrinsic to the photodetector. The object of this paper is fourfold. First, we calculate the variance of the output current in terms of parameters that characterize these two sources of randomness (the Fano factor for the light and the excess noise factor for the detector multiplication process). Second, we demonstrate that in the usual situation (Poisson photon arrivals), a simple relation between the output-current variance and the excess noise factor of the multiplication ensues. Third, we give explicit expressions for the excess noise factors of three photodetectors of interest: the conventional avalanche photodiode, the super-

lattice avalanche photodiode, and the photomultiplier tube. Finally, we compare and contrast the noise behavior of these three photodetectors, illustrating their relative merits.

The usual method for experimentally determining the excess noise factor for an APD involves a measurement of the variance of the output current[1] when the device is illuminated by a Poisson stream of photons [1]. Yet the theoretically calculated excess noise factor is defined in terms of the normalized second moment of the gain random variable when a single photocarrier initiates the multiplication [2]–[4]. The relationship between the two quantities is generally obtained individually for each APD [2]–[5]. Our first task is to derive a general formula that relates these quantities for an arbitrary source of light and for an arbitrary detector multiplication process. When the number of photons at the input to the detector is Poisson, the output-current variance turns out to be directly proportional to the excess noise factor.

Explicit formulas for the excess noise factor are presented for several special cases: the conventional avalanche photodiode (CAPD), the superlattice avalanche photodiode (SAPD), and the photomultiplier tube (PMT). The results for the CAPD are in accord with the expressions obtained earlier by McIntyre [2], [3], whereas the results for the single-carrier SAPD agree with those reported by Capasso et al. [5] for the graded-gap staircase APD. The expressions for the double-carrier SAPD are new, although they are related to expressions obtained by van Vliet et al. [6]. The formulas for the PMT were derived long ago by Zworykin et al. [7] and by Shockley and Pierce [8]. The noisiness of the three photodetectors is compared graphically.

The excess noise factor is a useful statistic because it represents, in a compact way, the lowest order statistical properties of the gain fluctuations that introduce multiplicative noise. However, it must be recognized that, aside from photon fluctuations, the excess noise factor does not provide a complete statistical description of the electron current. While it is useful for the calculation of quantities such as the conventional signal-to-noise ratio (SNR) for

[1] If only the ac or "signal" portion of the current is considered, the variance is equivalent to the mean square current.

analog detection, it is inadequate for describing the performance of a digital-signal information transmission system [9]. Instead, measures such as probability of detection and probability of error must be used for such systems. These latter quantities are strongly dependent on the tails of the counting distributions (beyond the second moment) and therefore require a more complete statistical description of the electron current (e.g., the counting distribution) [10]–[12].

II. THEORY

Consider a point process representing the primary (photon-generated) carriers. Let the number of these carriers generated within the time interval [0, T] be described by the discrete random variable a. Each of these primary carriers, in turn, is assumed to independently produce M daughter carriers (M is the discrete gain random variable representing the carrier multiplication). The total number of electrons n produced at the output of the device is the quantity of interest.

If a and M are statistically independent, which it is safe to assume, then

$$\langle n \rangle = \langle M \rangle \langle a \rangle \tag{1}$$

and

$$\text{Var} (n) = \langle M \rangle^2 \text{Var} (a) + \langle a \rangle \text{Var} (M). \tag{2}$$

The angular brackets $\langle \cdot \rangle$ represent the ensemble average and Var $(\cdot)$ represents the count variance. These relationships are known as the Burgess variance theorem [13]–[15]. Dividing (2) by (1) provides

$$\mathfrak{F}_n = \langle M \rangle \mathfrak{F}_a + \mathfrak{F}_M \tag{3}$$

where the Fano factors $\mathfrak{F}_j$ are defined as

$$\mathfrak{F}_j = \text{Var} (j)/\langle j \rangle, \quad j = n, a, M. \tag{4}$$

The quantity $\mathfrak{F}_a$ is the Fano factor for the photogenerated carriers. For many photon-counting distributions, including the Poisson, the Fano factor is invariant to random deletion [14], in which case $\mathfrak{F}_a$ is the same as the Fano factor for the photons incident on the device ($\mathfrak{F}_{\text{photon}}$). Equation (3) is are then independent of the detector quantum efficiency η.

In many detectors, the processes a and n are filtered (continuous) versions of their discrete counterparts. In that case, we use the spectral form of the Burgess variance theorem [15]

$$\langle I_n \rangle = \langle M \rangle \langle I_a \rangle \tag{5}$$

and

$$S_n = \langle M \rangle^2 S_a + 2q \langle I_a \rangle \text{Var} (M). \tag{6}$$

The quantities $\langle I_n \rangle$ and $\langle I_a \rangle$ and S_n and S_a represent the mean currents and power spectral densities, respectively, for the n and a processes; q is the electronic charge. Since the primary process satisfies

$$\text{Var} (I_a) = (q/T)^2 \text{Var} (a). \tag{7}$$

the relation $B = 1/2T$ (where B is the bandwidth of the system) can be used to obtain

$$S_a = 2q \langle I_a \rangle \mathfrak{F}_a, \tag{8a}$$

thereby a.lowing (6) to be simplified to

$$S_n = 2q \langle I_a \rangle [\langle M \rangle^2 \mathfrak{F}_a + \text{Var} (M)]. \tag{8b}$$

The excess noise factor F_e is defined as the normalized second moment of the gain random variable *for a single input photocarrier*, i.e.,

$$F_e = \langle M^2 \rangle/\langle M \rangle^2. \tag{9}$$

Equation (9) can be used in conjunction with (8) to provide

$$S_n = 2q \langle I_a \rangle \langle M \rangle^2 [\mathfrak{F}_a + (F_e - 1)] \tag{10a}$$

and

$$\text{Var} (I_n) = 2q \langle I_a \rangle B \langle M \rangle^2 [\mathfrak{F}_a + (F_e - 1)], \tag{10b}$$

which is the desired general relationship. Combining (9) with (2) leads to the discrete analog to (10b), which may be written as

$$\text{Var} (n) = \langle a \rangle \langle M \rangle^2 [\mathfrak{F}_a + (F_e - 1)] \tag{10c}$$

or equivalently as

$$\mathfrak{F}_n = \langle M \rangle [\mathfrak{F}_a + (F_e - 1)]. \tag{10d}$$

The excess noise factor can also be expressed in terms of the mean and variance of the gain by

$$F_e = 1 + [\text{Var} (M)/\langle M \rangle^2]. \tag{11}$$

For deterministic multiplication

$$\text{Var} (M) = 0, \quad F_e = 1 \tag{12}$$

whence the name "excess noise factor."

The laboratory measurement of an experimental excess noise factor ϕ_e is often carried out by determining the true ac mean-square current at the output of the APD under study and the true ac mean-square current at the output of a device identical in all respects except that with unity multiplication. The ratio of these two currents (which are, properly speaking, variances) provides the experimental excess noise factor ϕ_e [1]. We therefore have

$$\phi_e = \text{Var} (I_n)/\text{Var} (I_a) = S_n B/S_a B. \tag{13}$$

Using (10a) together with (8a) leads to

$$\phi_e = (\langle M \rangle^2/\mathfrak{F}_a) [\mathfrak{F}_a + (F_e - 1)]. \tag{14a}$$

Actual experimental measurements of ϕ_e are invariably carried out with a source of radiation that generates a Poisson flow of photocarriers. In that case, $\mathfrak{F}_a = 1$ so that (14a) becomes

$$\phi_e = \langle M \rangle^2 F_e. \tag{14b}$$

This explicitly demonstrates the proportionality of the experimentally determined and theoretically calculated excess noise factors for Poisson light.

The general result for ϕ_e given in (14a) may be equiv-

alently expressed in terms of the mean $\langle M \rangle$ and Fano factors $\mathfrak{F}_M$ and $\mathfrak{F}_a$. Using (11), we obtain

$$\phi_c = \langle M \rangle (\langle M \rangle + \mathfrak{F}_M/\mathfrak{F}_a). \tag{15a}$$

In the case of Poisson photocarriers, this simplifies to

$$\phi_c = \langle M \rangle (\langle M \rangle + \mathfrak{F}_M). \tag{15b}$$

III. Excess Noise Factor for the Conventional APD

An expression for the excess noise factor F_c associated with avalanching in a uniformly multiplying p-n junction was first derived by McIntyre [2]. When only electrons are injected into the depletion layer, but both electrons and holes have the capability to impact ionize, the result can be written as

F_c (electron injection)

$$= [k_c \langle M \rangle + (2 - 1/\langle M \rangle)(1 - k_c)]; \tag{16a}$$

when only holes are injected, it becomes

F_c (hole injection)

$$= [\langle M \rangle/k_c - (2 - 1/\langle M \rangle)(1 - k_c)/k_c]. \tag{16b}$$

Equations (16a) and (16b) are, as indicated above, valid for single-carrier-initiated/double-carrier multiplication (SCIDCM). The quantity

$$k_c = \beta/\alpha \tag{17}$$

in (16a) and (16b) represents the ratio of hole-ionization probability per unit length β to electron-ionization probability per unit length α. This ratio is assumed to be independent of the electric field $\vec{E}$ and constant throughout the avalanche region.² Equation (16b) may be obtained directly from (16a) by using the substitution $k_c \rightarrow 1/k_c$. If electrons and holes are both injected, the overall result is obtained by adding the two partial results.³ The lower the value of k_c, the lower the device noise. In experimentally determining the APD excess noise factor, the quantity that is directly measured is the variance of the output current in response to a Poisson stream of photons at the input, as specified in (10b) with $\mathfrak{F}_a = 1$. Using (16a) and (16b) in (10b) provides results that accord with those derived by McIntyre [2].

The average multiplication (mean gain) $\langle M \rangle$ for a CAPD with pure electron injection, expressed as a function of the distance from the edge of the depletion layer x, is [2]–[4]

$$\langle M \rangle = (1 - k_c)/\{\exp[\alpha(k_c - 1)x] - k_c\}. \quad k_c \neq 1 \tag{18}$$

---

²McIntyre [3] demonstrated that even if β is not proportional to α, a suitable value for k, called k_{eff}, can be defined for F, if $\langle M \rangle$ is large. Furthermore, additional noise is introduced when light is absorbed on both sides of the junction, so that both electrons and holes are injected into the avalanche region (this is double-carrier-initiated-double-carrier multiplication or DCIDCM). In that case, an effective excess noise factor F_{e*} can be defined [4]. McIntyre's theoretical results for Si APD's were experimentally verified by Conradi [1].

The performance of digital-communication systems incorporating CAPD's was examined in detail by Personick. His initial treatment of this topic [10] dealt with multiplication involving a single ionizing carrier, as well as multiplication involving two carriers with equal ionization coefficients. In a subsequent generalization [11], he obtained upper-bound performance results for double-carrier devices with unequal ionization coefficients. Personick's results are consistent with those obtained by McIntyre [3].

An important special case is that of the CAPD under single-carrier-initiated/single-carrier multiplication (SCISCM) conditions. This provides the lowest possible noise. Setting $k_c = 0$ in (16a) or $k_c = \infty$ in (16b) leads to

$$F_c = 2 - 1/\langle M \rangle \tag{19}$$

which, with the help of (10b) with $\mathfrak{F}_a = 1$, gives rise to

$$\text{Var} (I_n) = 2q \langle I_a \rangle B \langle M \rangle [2\langle M \rangle - 1]. \tag{20}$$

The average multiplication, readily obtained from (18), is then

$$\langle M \rangle = \exp(\alpha x). \tag{21}$$

Inserting (21) into (20) gives the expression for the variance of a filtered Yule–Furry birth process with a Poisson initial population [16]. The identity between the statistics of the SCISCM CAPD and the Yule–Furry process is confirmed by the gain distribution (in response to a single initiating event); it is the shifted Bose–Einstein distribution in both cases [10], [12], [16].

In the case where the ionization coefficients for electrons and holes are equal ($\alpha = \beta$; $k_c = 1$), (16a) and (16b) become

$$F_c = \langle M \rangle. \tag{22}$$

whereupon (10b), with $\mathfrak{F}_a = 1$, provides

$$\text{Var} (I_n) = 2q \langle I_a \rangle B \langle M \rangle^3. \tag{23}$$

as first obtained by Tager [17]. The average multiplication is then

$$\langle M \rangle = 1/(1 - \alpha x), \quad k_c = 1 \tag{24}$$

For a device operated with either pure-electron or pure-hole delta-function injection (i.e., pure injection at one or the other edge of the depletion layer), this is the noisiest situation. However, for $\alpha \neq \beta$, F_c can be even greater than or less than $\langle M \rangle$, depending on where the light is absorbed in the junction.³

IV. Excess Noise Factor for the Superlattice APD

An illustrative example of a superlattice APD (SAPD) is the staircase avalanche photodiode. It is a graded-gap multilayer device proposed by Williams, Capasso, and Tsang [5], [18], [19] for low-noise light detection in the near-infrared region of the spectrum. The device is of interest for fiber-optic communications. It is designed to achieve an enhancement of the impact-ionization proba-

bility ratio, thereby minimizing the hole-electron feedback noise associated with conventional III–V quaternary APD's for which $k_c \approx 1$ [20]. Furthermore, because the electron multiplication can occur only at a finite number of discrete locations in the device, the variability of the number of electrons generated per detected photon is also reduced relative to the CAPD. (It has already been experimentally shown that the first superlattice APD structure, proposed by Chin *et al.*, can provide an enhanced ionization ratio [21]–[25].) The fabrication of a staircase device in this region of the spectrum, using molecular-beam epitaxy, is currently underway at AT&T Bell Laboratories [24].

Although we deal with the graded-gap staircase SAPD extensively for purposes of illustration, the analysis pre-

their model a good starting point for describing the excess noise factor for the double-carrier instantaneous-multiplication SAPD. Although the theory is appropriate as it stands for those SAPD's in which both electron and hole ionizations occur at discrete locations (e.g., the multiquantum-well SAPD [23]), it must be modified for the staircase SAPD. In this latter case, we must incorporate a proper *continuous* theory for the hole-ionization probability [5] into the discrete shifted-Bernoulli theory of van Vliet *et al*.

The excess noise factor F_e for the double-carrier SAPD with electron injection may then be obtained from the expression for the variance derived by van Vliet *et al*. [6, eq. (73)] using the notational replacements $M_N \rightarrow \langle M \rangle$, $\lambda \rightarrow P$, and $k \rightarrow k_s$ along with (11). This yields

$$F_e \text{ (electron injection)} = 1 + \frac{(1 - 1/\langle M \rangle)(1 - k_s)}{2 + P(1 + k_s)}$$

$$\times \left[-P + 2\frac{1 - k_s P^2}{1 + k_s P} \langle M \rangle k_s \frac{1 + P}{1 - k_s} + \frac{1}{1 + P} \right] \tag{25}$$

sented here is applicable for any SAPD in which the carrier transport is perpendicular to the superlattice planes. In such structures, the carriers encounter a potential discontinuity at the heterointerfaces at each period of the multilayer structure. Thus, our results will also apply to the multiquantum-well SAPD structure of Chin *et al* [21]–[23], [25], the doped-quantum-well SAPD structure of Blauvelt *et al* [26], and the stored-carrier multiquantum-well SAPD [27]–[29]. However, the results will not apply to the channeling APD [30], [31], nor to other devices in which the carriers are spatially separated by means of a transverse field with transport taking place in the plane of the layers.

The gain, excess noise factor, and electron probability distribution at the output of a staircase SAPD have recently been calculated as a function of the number of stages of the device m and the electron impact-ionization probability per stage P under the SCISCM assumption [32].[^1] This analysis is valid not only for the staircase SAPD but for any of the perpendicular-carrier-transport SAPD's.

We now extend these results to allow for residual hole-initiated ionization in the graded regions of the device arising from the applied electric field. (The valence-band steps are of the wrong sign to assist hole-initiated ionization, indeed they may lead to hole trapping. Also, because of the opposing conduction-band quasi-electric field, the electrons can impact ionize only at the conduction-band discontinuities.) The mathematical results follow from the treatment provided by van Vliet *et al* [6], with appropriate extension and reinterpretation. The model provided by these authors was intended to describe multiplication noise in CAPD's. However, their use of a deterministic number of shifted-Bernoulli stages makes

As previously, P is the electron impact-ionization probability per stage and $\langle M \rangle$ is the average overall multiplication for the device. The ratio of the hole-ionization probability per stage Q to the electron-ionization probability per stage P defines k_s for the SAPD, i.e.,

$$k_s = Q/P. \tag{26a}$$

For the staircase SAPD, Q is given by [5]

$$Q = \exp\left(\int_0^L \beta \, dz\right) - 1 \tag{26b}$$

where β is the hole-ionization coefficient (probability per unit length) in the graded region and L is the length of each stage.

The average multiplication $\langle M \rangle$ is obtained with the help of [6, eq. (53)] and (26a) which, together with the notational substitutions $\lambda \rightarrow P$, $\mu \rightarrow Q$, $N \rightarrow m$, yield

$$\langle M \rangle = \frac{(1 + P)^m (1 - k_s)}{(1 + k_s P)^m - k_s (1 + P)^{m+1}} \tag{27}$$

From (27), it is apparent that the expression for the excess noise factor in (25) could readily be expressed as a function of P, k_s, and m instead of P, k_s, and $\langle M \rangle$. When $k_s \neq 0$, the average multiplication in (27) will increase without limit for certain parameter values (i.e., avalanche breakdown will occur). The validity of (27) is restricted to the parameter space below avalanche breakdown.

The modified excess noise factor $(F_e - 1)$ for the SCIDCM SAPD is plotted as a function of the average multiplication $\langle M \rangle$ with the help of (25) and (27). The modified excess noise factor is used because it is conveniently displayed on double-logarithmic coordinates. Furthermore, as is evident from Section II, it is the pertinent measure in the absence of quantum fluctuations. In Figs. 1 and 2, we illustrate its behavior for $m = 3$ and 10, respectively. Results for several values of k are presented

[^1]: The impulse response function of a SAPD is also affected by incorporating the effects of randomness in the carrier multiplication process.
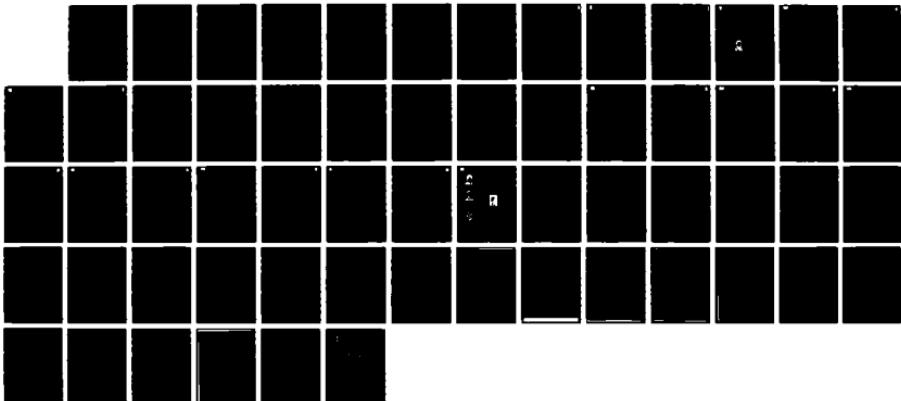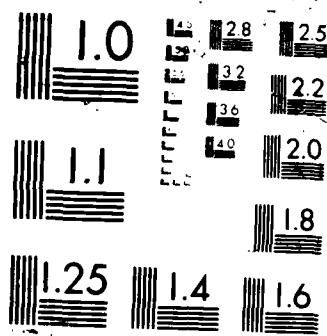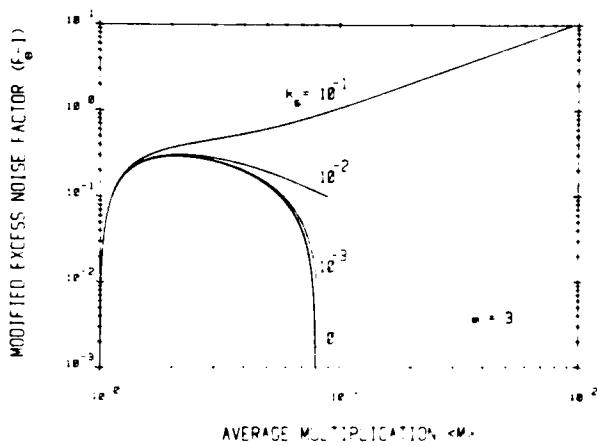
Fig. 1. Modified excess noise factor $F_e - 1$ versus average multiplication $\langle M \rangle$ for the single-carrier-initiated/double-carrier multiplication (SCIDCM) SAPD. The modified excess noise factor is plotted because it can be conveniently displayed on double-logarithmic coordinates. It is the pertinent measure in the absence of quantum fluctuations. The number of stages $m = 3$. The behavior for different values of k_s is illustrated parametrically. The curve for $k_s = 0$ corresponds to the single-carrier-initiated/single-carrier multiplication (SCISCM) SAPD. It is apparent that even small deviations of k_s from 0 result in substantial excess noise.
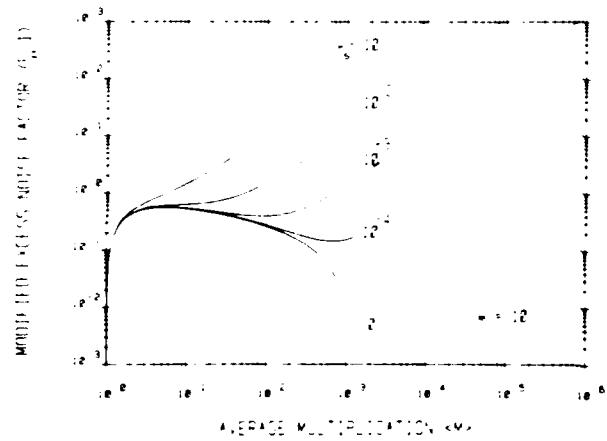

Fig 2. Modified excess noise factor $F_e - 1$ versus average multiplication $\langle M \rangle$ for the SCIDCM SAPD with $m = 10$. Again, the curve $k_s = 0$ corresponds to the SCISCM SAPD. Comparison to Fig. 1 demonstrates that as m increases, the deleterious effects of residual hole ionization become more pronounced.

parametrically. The $k_s = 0$ curve corresponds to the SCISCM SAPD. As for the CAPD, the lower the value of k_s, the lower the noise. It is apparent that even small deviations of k_s from 0 result in substantial excess noise. This effect is more pronounced as m increases. These results provide limits on the residual hole ionization that is tolerable in an SAPD. Based on a many-particle Monte Carlo simulation, Brennan [33] has recently estimated k_s to be $\approx 10^{-1}$ for the multiquantum-well SAPD of Chin et al. [23] and $\approx 10^{-2}$ for the staircase SAPD of Williams et al. [18], in the GaAs/AlGaAs system. He has also shown that even lower values of k_s may be achievable by using the doped-quantum-well SAPD structure of Blauvelt et al. [26].

The most important special case, of course, is that of the SCISCM device, which is the lowest noise SAPD. From (27), the average multiplication for the single-carrier multiplication SAPD is easily shown to be

$$\langle M \rangle = (1 + P)^m. \tag{28}$$

The excess noise factor is obtained by setting $k_s = 0$ in (25), which leads to

$$F_e = 1 + [(1 - P)/(1 + P)][1 - (1 + P)^{-m}] \tag{29a}$$

$$= \langle M \rangle^{-1} + 2\langle M \rangle^{-1/m} - 2\langle M \rangle^{-1-1/m}. \tag{29b}$$

Using (14b), the experimental excess noise factor ϕ_e will then be given by

$$\phi_e = (1 + P)^{2m} + [(1 - P)/(1 + P)][(1 + P)^{2m}$$
$$- (1 + P)^m] \tag{30}$$

which, with the help of (10b) and $\mathfrak{F}_a = 1$, corresponds to the output current variance

$$\text{Var}(I_n) = 2q\langle I_a \rangle B\{(1 + P)^{2m} + [(1 - P)/(1 + P)]$$
$$\cdot [(1 + P)^{2m} - (1 + P)^m]\}. \tag{31}$$

Equation (29a) agrees with the formula obtained by Capasso et al. [5, eq. (2)]. Equation (31) is also in accord with their result [5, eq. (1)], provided that the quantity $\langle i^2 \rangle$ in [5, eq. (1)] is interpreted as the power spectral density S_n. All of the formulas presented here are in agreement with those reported in [32]. Equation (29b) is displayed in Figs. 1 and 2 ($k_s = 0$).

Carrying (29b) to the limit $m \to \infty$ leads to the result

$$F_e = 2 - 1/\langle M \rangle, \tag{32}$$

which is identical to that given for the SCISCM CAPD in (19). This is as expected; in this limit, there is an infinite number of stages and the probability is vanishingly small that a carrier is produced by impact ionization in any one given stage of the device.

Finally, we consider the case of equal ionization coefficients for electrons and holes in the SAPD ($P = Q$; $k_s = 1$). The excess noise factor, output-current variance, and average multiplication then become

$$F_e = \langle M \rangle - (\langle M \rangle - 1)^2/m\langle M \rangle \tag{33}$$

$$\text{Var}(I_n) = 2q\langle I_a \rangle B[\langle M \rangle^3 - \langle M \rangle(\langle M \rangle - 1)^2/m] \tag{34}$$

and

$$\langle M \rangle = 1/(1 - Pm), \tag{35}$$

respectively. Carrying (33) to the limit $m \to \infty$ provides

$$F_e = \langle M \rangle \tag{36}$$

which is identical to the SCIDCM CAPD result given in (22) for the same reasons as indicated above.
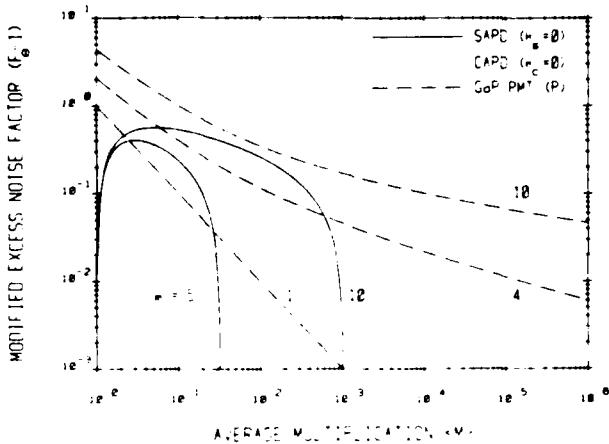
Fig 3 Modified excess noise factor $F_e - 1$ versus average multiplication $\langle M \rangle$ for the SCISCM SAPD (solid curves) with $m = 5$ and 10, the SCISCM CAPD (dotted curve), and the high-gain GaP-first-dynode PMT with Poisson multiplication and $A = 10$ (dashed curves) for $m = 1, 4, 10$. This figure illustrates the best possible behavior for all three devices. Although the theoretical excess noise factor of the SAPD is always superior to that of the CAPD, the differential cannot be large because $F_e < 2$ for both devices. The PMT exhibits high-gain, low-noise behavior for all useful values of $\langle M \rangle$. In terms of the excess noise factor, it can exhibit better performance than the SAPD.
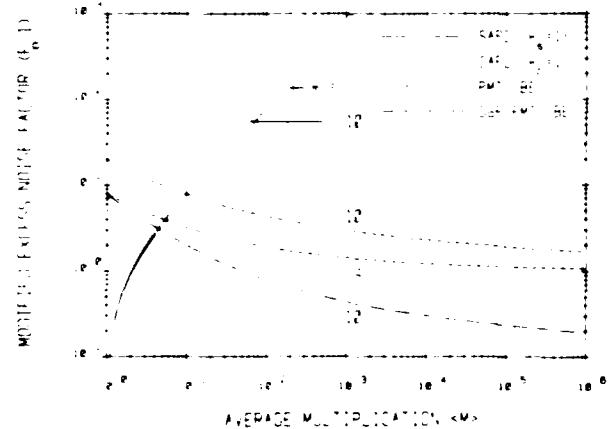

Fig 4 Modified excess noise factor $F_e - 1$ versus average multiplication $\langle M \rangle$ for the SCIDCM SAPD ($k_s = 1$, solid curves) with $m = 5$ and 10, the SCIDCM CAPD ($k_c = 1$, dotted curve), the identical-dynode PMT with Bose-Einstein multiplication (short-dash curves) and with $m = 4$ and 10 and the high-gain GaP-first-dynode 10-stage PMT with Bose-Einstein multiplication and with $A = 10$ (long-dash curve). This figure illustrates the worst possible behavior for all three devices, assuming single carrier (delta-function) injection. For the CAPD, F_e is identically equal to $\langle M \rangle$, whereas for the SAPD, F_e increases approximately as $\langle M \rangle$. These results illustrate explicitly that the feedback associated with double-carrier multiplication in APD's gives rise to far more noise than does uncertainty in the ionization locations. The quintessentially single-carrier PMT displays far and away the lowest values of the excess noise factor.

V. COMPARISON OF EXCESS NOISE FACTORS FOR THE SAPD AND THE CAPD

The modified excess noise factors for the SAPD (solid curves) and CAPD (dotted curves) are plotted versus the average multiplication $\langle M \rangle$ in Figs. 3 and 4 for $k_s = k_c = 0$ and $k_s = k_c = 1$, respectively. The results in Fig. 3 are for SCISCM conditions, representing the best possible

behavior of both devices. The excess noise factors for both the SAPD and the CAPD then always lie below 2, as is apparent from (19) and (29a). Although the theoretical performance of the SAPD is always superior to that of the CAPD, the differential (in terms of excess noise factor) cannot be very large since $F_e < 2$ for both cases.

In the opposite limit (SCIDCM with equal impact-ionization probabilities), the results are displayed in Fig. 4. This is the worst possible behavior for both devices, assuming single-carrier (delta-function) injection. F_e for the CAPD is identically $\langle M \rangle$ from (22). F_e for the SAPD also increases in approximate proportion to $\langle M \rangle$, but with the coefficient $(1 - 1/m)$, which is slightly below unity, as can be discerned from (33). Thus, the SAPD excess-noise curves lie slightly below the CAPD curve for finite m, coming ever closer as $m \to \infty$, in accord with (36).

The results presented in Figs. 3 and 4 illustrate explicitly that the feedback noise introduced by the process of double-carrier multiplication is far more deleterious to noise performance than is the randomness associated with uncertainty in the locations of the ionizations.

VI. EXCESS NOISE FACTOR FOR THE PHOTOMULTIPLIER TUBE

The photomultiplier tube (PMT) is one of the oldest and most versatile of light detectors, having been developed about 1935 [7]. From a noise point of view, it has the distinct advantage of being a single-carrier device (SCISCM) since the electrons travel in vacuum. An excellent description of essentially all aspects of PMT operation is available in the *RCA Photomultiplier Handbook* [34]; in particular, the reader is directed to Appendix G (pp. 160–176) for a comprehensive discussion of the statistical theory of PMT noise.

The average multiplication for an m-stage device is given by [34, eq. (G-55)]

$$\langle M \rangle = \prod_{k=1}^{m} \langle \delta_k \rangle \qquad (37)$$

where $\langle \delta_k \rangle$ is the mean secondary-emission gain at the kth stage. The random variable δ_k represents the kth-stage secondary-emission gain. Using an expression for the variance of the multiplication M at the output of the PMT [34, eq. (G-56)] together with (11) gives the excess noise factor

$$F_e = 1 + \frac{\text{Var}(\delta_1)}{\langle \delta_1 \rangle^2} + \frac{\text{Var}(\delta_2)}{\langle \delta_1 \rangle \langle \delta_2 \rangle^2} + \cdots$$

$$+ \frac{\text{Var}(\delta_m)}{\langle \delta_1 \rangle \langle \delta_2 \rangle \cdots \langle \delta_{m-1} \rangle \langle \delta_m \rangle^2}. \qquad (38)$$

This expression is of general validity for single-carrier discrete multiplication processes. (Thus, it can also be used for the SCISCM SAPD when the individual stages have different values of P.) It, and variants of it, were obtained early on [7], [8], [35]. It is apparent from the sequence of denominators in (38) that the gain of the first

stage $\langle \delta_1 \rangle$ has a substantial influence on F_e: the higher this gain, the lower the contribution to the excess noise from the subsequent stages. It is this mathematical property that spurred the development of high-gain GaP-first-dynode PMT's at RCA [36].

Two useful special cases of (37) and (38) involve PMT's with identical dynodes and PMT's with a high-gain first dynode. We first assume that all m dynodes of the device are equivalent, so that the δ_k are independent and identically distributed (iid) random variables for all k, with mean and variance given by $\langle \delta \rangle$ and Var (δ), respectively. Then (37) becomes

$$\langle M \rangle = \langle \delta \rangle^m \qquad (39)$$

whereas (38) becomes

$$F_e = 1 + \frac{\text{Var } (\delta)}{\langle \delta \rangle (\langle \delta \rangle - 1)} \left[1 - \frac{1}{\langle \delta \rangle^m} \right]. \qquad (40)$$

These expressions are exact and applicable to any single-carrier device with m identical stages. Thus, the formula for the SCISCM SAPD excess noise factor given in (29a) may be obtained by setting $\langle \delta \rangle = 1 + P$ and Var $(\delta) = P(1 - P)$.

We next assume that all stages produce iid secondary electrons, with mean and variance $\langle \delta \rangle$ and Var (δ) respectively, except for the first stage for which

$$\langle \delta_1 \rangle = A \langle \delta \rangle \qquad (41a)$$

and

$$\text{Var } (\delta_1) = A \text{ Var } (\delta) \qquad (41b)$$

where A is a constant. When $A \gg 1$, this characterizes the high-gain first-dynode PMT (e.g., GaP). The mean gain (37) then becomes

$$\langle M \rangle = A \langle \delta \rangle^m \qquad (41c)$$

whereas the excess noise factor (38) becomes

$$F_e = 1 + \frac{\text{Var } (\delta)}{A \langle \delta \rangle (\langle \delta \rangle - 1)} \left[1 - \frac{1}{\langle \delta \rangle^m} \right]. \qquad (42)$$

The similarity between (42) and (40) is obvious, with the prefactor $1/A$ on the right-hand side of (42) succinctly representing the excess-noise-factor advantage of the high-gain first-dynode tube.

To proceed further, the variance of the gain at the various stages must be specified. This is determined by the secondary-emission process at the dynodes. The simplest and most frequently used model invokes Poisson secondary-emission multiplication at every stage (Var $(\delta) = \langle \delta \rangle$) [7], [8], [35], [37]–[39]. A more versatile model is provided by the neagative-binomial (or Polya) distribution, which has been used by Prescott [40] and others [34], [41], [42]. This distribution arises from a mixture of Poisson distributions whose means are smeared in accordance with the gamma distribution [43]. Physically, the smearing is thought to arise from the variability of the secondary-emission efficiency across the surface of the dynode

[34], [40]. In this case,

$$\text{Var } (\delta) = \langle \delta \rangle + \langle \delta \rangle^2/D \qquad (43)$$

where D is the "degrees-of-freedom" parameter describing the extent of the smearing [44]. Two special limits of the negative binomial are the Poisson distribution for which $D = \infty$ (the least noisy) and the Bose–Einstein (or Furry) distribution for which $D = 1$ (the most noisy).

Of the various possibilities implicit in the results of this section, the overall lowest excess noise factor obtains for a high-gain first-dynode PMT with Poisson multiplication. In that case, using (41)–(43), we obtain

$$F_e = 1 + \frac{1}{\langle M \rangle} \left[\frac{\langle M \rangle/A - 1}{(\langle M \rangle/A)^{1/m} - 1} \right]. \qquad (44)$$

Equation (44) is plotted as the dashed curves in Fig. 3 for $A = 10$ ($m = 1, 4, 10$). The gain of the $m = 1$ (4) GaP PMT is approximately the same as that of the $m = 5$ (10) SAPD. It is apparent that the excess noise factor of the PMT can fall below that of the SAPD. Of course, the overall gain of a PMT can stretch to $\approx 10^6$, which is far and away greater than that achievable by any APD.

The excess noise factor for Poisson secondary-emission multiplication without the benefit of the high-gain first dynode (all dynodes identical) is obtained by setting $A = 1$ in (44). When $\langle M \rangle \gg 1$, F_e then takes the well-known approximate form

$$F_e \approx \langle M \rangle^{1/m}/(\langle M \rangle^{1/m} - 1) = \langle \delta \rangle/(\langle \delta \rangle - 1), \qquad (45)$$

signifying essentially noise-free multiplication even in this case.

The excess noise factor for a high-gain first-dynode PMT with Bose–Einstein secondary-electron statistics is obtained from (41)–(43) with $D = 1$. It is

$$F_e = 1 + \frac{1 + (\langle M \rangle/A)^{1/m}}{\langle M \rangle} \left[\frac{\langle M \rangle/A - 1}{(\langle M \rangle/A)^{1/m} - 1} \right]. \qquad (46)$$

which is plotted as the long-dash curve in Fig. 4 for $A = 10$ ($m = 10$).

Finally, the noisiest of the PMT cases considered here arises for Bose–Einstein secondary-electron statistics with all dynodes equivalent. This result is obtained by setting $A = 1$ in (46). For $\langle M \rangle \gg 1$, F_e then takes the approximate form

$$F_e \approx 2 \langle M \rangle^{1/m}/(\langle M \rangle^{1/m} - 1)$$
$$= 2 \langle \delta \rangle/(\langle \delta \rangle - 1). \qquad (47)$$

This is only a factor of 2 greater than the excess noise factor in (45) for Poisson multiplication. This case is plotted as the short-dash curves in Fig. 4 ($m = 4, 10$).

It is apparent that, even at its noisiest, the PMT is a relatively quiet device. This is a consequence of its single-carrier vacuum character. The low excess noise factor has made the PMT an indispensible tool in optics laboratories since the 1930's and it is not likely to be relegated to the junk heap any time soon.

VII. CONCLUSION

We have obtained a general expression (10b) for the output-current variance of a photodetector in terms of the Fano factor of the photoelectrons $\mathfrak{F}_a$ and the excess noise factor of the multiplication process intrinsic to the detector F_e. Sources of noise such as dark current, Johnson noise, and $1/f$ noise were not included in the simple formulation considered here; they can be easily incorporated, however. The analysis assumed instantaneous multiplication in that the measurement time was assumed to be greater than the time response of the signal [32].

The variance can be cast in terms of a current signal-to-noise ratio at the output of the photodetector given by

$$SNR\ (I_n) = \langle I_n \rangle / [\text{Var}\ (I_n)]^{1/2}$$

$$= [\langle I_a \rangle / 2qB(\mathfrak{F}_a + F_e - 1)]^{1/2} \quad (48)$$

where $\langle I_a \rangle = \eta \langle I_{photon} \rangle$. In the usual situation of Poisson photon arrivals ($\mathfrak{F}_a = 1$), the output-current variance is proportional to F_e in accordance with

$$\text{Var}\ (I_n) = 2q\langle I_a \rangle\ B\langle M \rangle^2 F_e \quad (49)$$

whereupon the experimental excess noise factor is

$$\phi_e = \langle M \rangle^2 F_e \quad (50)$$

and the SNR is

$$SNR\ (I_n) = [\langle I_a \rangle / 2qBF_e]^{1/2}. \quad (51)$$

In this case, optimization of the SNR simply involves minimizing F_e, independent of $\langle M \rangle$. In the presence of thermal noise and/or dark noise, however, optimization of the SNR is achieved at specific values of $\langle M \rangle$ and F_e.

Under conditions of Poisson photon excitation (or indeed for any distribution of photons for which $\mathfrak{F}_a$ is constant and independent of $\langle a \rangle$), the right-hand side of (3) depends only on the random gain M. Thus, the ratio of overall count variance to overall count mean is constant and independent of the excitation level. The output current from any multiplying detector illuminated by such light therefore has shot-noise-like behavior. A useful consequence of this property is the applicability of the square-root normalizing transformation [45]. With the help of this computational tool, system performance can be readily evaluated in approximate form [46]. The presence of additive thermal noise and/or dark noise will, however, destroy this shot-noise-like behavior.

In the complete absence of photoelectron fluctuations ($\mathfrak{F}_a = 0$), and thermal and dark noise, the output-current variance is no longer proportional to F_e, but is instead proportional to $(F_e - 1)$. Thus, Var $(I_n) \rightarrow 0$ as $F_e \rightarrow 1$. In this idealized case, the signal-to-noise ratio at the output of the photodetector is

$$SNR\ (I_n) = [\langle I_a \rangle / 2qB(F_e - 1)]^{1/2}, \quad (52)$$

which can become large for a photodetector with $F_e \rightarrow 1$.

Expressions for the excess noise factors for CAPD, SAPD, and PMT photodetectors have been set forth and graphically displayed. In terms of existing devices, small Si CAPD's with high quantum efficiency and near-ideal performance ($F_e = 2.6$ corresponding to $k_c = 0.006$ at $\langle M \rangle = 100$) have been fabricated in the wavelength region $0.4 < \lambda < 0.95\ \mu m$ [47]. Devices that are even more quiet, with $F_e < 2.2$ corresponding to $k_c < 0.002$ at $\langle M \rangle = 100$, appear to be possible [48]. CAPD devices with essentially SCISCM properties are therefore currently available at this wavelength. However, quaternary devices are generally used in the wavelength region $\lambda \approx 1.5\ \mu m$. Unfortunately, these have $k_c \approx 1$ so that F_e is much higher [20]. Dark current and leakage current may also present difficulties in such devices.

SAPD's offer promise in this longer wavelength region. These devices have potential as small, high quantum efficiency, low-voltage photodetectors, with low leakage current [5], [20]. Although SAPD's can, in principle, exhibit minimal noise ($F_e \rightarrow 1$), their average multiplication is restricted ($\langle M \rangle \leq 2^m$ for SCISCM operation). Residual hole ionization is also a potential problem and it is imperative to construct devices in which this effect is minimized [33]. The dark-current behavior of various SAPD's has not yet been established. Attempts are currently underway to construct a long-wavelength staircase SAPD [24].

Photomultiplier tubes have remarkably low excess noise, along with the desirable properties of low dark current, high gain, good pulse resolution, and ease of operation in the photon-counting mode. However, as is well known, they suffer from limited quantum efficiency, large size, high-voltage requirements, luminescence noise, and afterpulsing due to H^+ ions or inverse photoemission [34].

PMT's are sometimes used to discriminate between single- and multiple-photoelectron events. This capability follows from the high gain as well as the narrowness of the multiplied electron distribution [34] and is associated with a minimal value of the gain Fano factor $\mathfrak{F}_M$. Such discrimination is more difficult for the CAPD which suffers from a (rather broad) shifted Bose-Einstein gain distribution. It may also be difficult for the SAPD which exhibits multiple peaks in the gain distribution. unless P is close to unity [32]. A related matter is the operation of a photodetector as a high-speed photon counter. This generally requires some 10^4 electrons/photon to overcome preamplifier Johnson noise [48]. Although this is readily achieved with a PMT, such gains are not easily attained with APD's. In particular, reaching this gain in a SCISCM SAPD would require a structure of some 15 stages [48].

As a final note, we reiterate that the excess noise factor is an inadequate measure for describing the performance of a digital-signal information-transmission system. Appropriate performance measures for such systems (e.g., probability of error) require knowledge of the electron-counting distributions. In particular, digital-system performance is strongly dependent on the tails of these distributions. Certain of the photodetectors discussed here will have more favorable shapes for minimizing error probabilities than will others. This will be elucidated in a

companion study of error probabilities for a simple optical receiver, to be presented elsewhere [12].

ACKNOWLEDGMENT

We are grateful to R. J. McIntyre, F. Capasso, K. Brennan, and J. Conradi for valuable discussions.

REFERENCES

[1] J. Conradi, "The distribution of gains in uniformly multiplying avalanche photodiodes: Experimental," *IEEE Trans. Electron Devices*, vol. ED-19, pp. 713-718, 1972.

[2] R. J. McIntyre, "Multiplication noise in uniform avalanche diodes," *IEEE Trans. Electron Devices*, vol. ED-13, pp. 164-168, 1966.

[3] ——, "The distribution of gains in uniformly multiplying avalanche photodiodes: Theory," *IEEE Trans. Electron Devices*, vol. ED-19, pp. 703-713, 1972.

[4] P. P. Webb, R. J. McIntyre, and J. Conradi, "Properties of avalanche photodiodes," *RCA Rev.*, vol. 35, pp. 234-278, 1974.

[5] F. Capasso, W. T. Tsang, and G. F. Williams, "Staircase solid-state photomultipliers and avalanche photodiodes with enhanced ionization rates ratio," *IEEE Trans. Electron Devices*, vol. ED-30, pp. 381-390, 1983.

[6] K. M. van Vliet, A. Friedmann, and L. M. Rucker, "Theory of carrier multiplication and noise in avalanche devices—Part II Two-carrier processes," *IEEE Trans. Electron Devices*, vol. ED-26, pp. 752-764, 1979.

[7] V. K. Zworykin, G. A. Morton, and L. Malter, "The secondary emission multiplier—A new electronic device," *Proc. IRE*, vol. 24, pp. 351-375, 1936.

[8] W. Shockley and J. R. Pierce, "A theory of noise for electron multipliers," *Proc. IRE*, vol. 26, pp. 321-332, 1938.

[9] G. E. Stillman and C. M. Wolfe, "Avalanche photodiodes," in *Semiconductors and Semimetals, Vol. 12, Infrared Detectors II*, R. K. Willardson and A. C. Beer, Eds. New York: Academic, 1977, pp. 291-393.

[10] S. D. Personick, "New results on avalanche multiplication statistics with application to optical detection," *Bell Syst. Tech. J.*, vol. 50, pp. 167-189, 1971.

[11] ——, "Statistics of a general class of avalanche detectors with applications to optical communication," *Bell Syst. Tech. J.*, vol. 50, pp. 3075-3095, 1971.

[12] M. C. Teich, K. Matsuo, and B. E. A. Saleh, "Counting distributions and error probabilities for optical receivers incorporating superlattice avalanche photodiodes," *IEEE Trans. Electron Devices*, to be published.

[13] R. E. Burgess, "Some topics in the fluctuation of photo-processes in solids," *J. Phys. Chem. Solids*, vol. 22, pp. 371-377, 1961.

[14] M. C. Teich and B. E. A. Saleh, "Effects of random deletion and additive noise on bunched and antibunched photon-counting statistics," *Opt. Lett.*, vol. 7, pp. 365-367, 1982.

[15] K. M. van Vliet and L. M. Rucker, "Noise associated with reduction, multiplication and branching processes," *Physica*, vol. 95A, pp. 117-140, 1979.

[16] K. Matsuo, M. C. Teich, and B. E. A. Saleh, "Poisson branching point processes," *J. Math. Phys.*, vol. 25, pp. 2174-2185, 1984.

[17] A. S. Tager, "Current fluctuations in a semiconductor (dielectric) under the conditions of impact ionization and avalanche breakdown," *Sov. Phys.—Solid State*, vol. 6, pp. 1919-1925, 1965.

[18] G. F. Williams, F. Capasso, and W. T. Tsang, "The graded bandgap multilayer avalanche photodiode: A new low-noise detector," *IEEE Electron Device Lett.*, vol. EDL-3, pp. 71-73, 1982.

[19] F. Capasso, "Band-gap engineering via graded gap, superlattice, and periodic doping structures: Applications to novel photodetectors and other devices," *J. Vac. Sci. Technol. B*, ser. 2, vol. 1, pp. 457-461, 1983.

[20] G. E. Stillman, V. M. Robbins, and N. Tabatabaie, "III-V compound semiconductor devices: Optical detectors," *IEEE Trans. Electron Devices*, vol. ED-31, pp. 1643-1655, 1984.

[21] F. Capasso, W. T. Tsang, A. L. Hutchinson, and G. F. Williams, "Enhancement of electron impact ionization in a superlattice A new avalanche photodiode with a large ionization rate ratio," *Appl. Phys. Lett.*, vol. 40, pp. 38-40, 1982.

[22] K. Mohammed, F. Capasso, J. Allam, A. Y. Cho, and A. L. Hutchinson, "New high-speed long-wavelength $Al_{0.48}In_{0.52}As$-

[23] R. Chin, N. Holonyak, G. E. Stillman, J. Y. Tang, and K. Hess, "Impact ionization in multilayered heterojunction structures," *Electron. Lett.*, vol. 16, pp. 467-469, 1980.

[24] F. Capasso, AT&T Bell Lab., Murray Hill, NJ, private communication.

[25] F.-Y. Juang, U. Das, Y. Nashimoto, and P. K. Bhattacharya, "Electron and hole impact ionization coefficients in GaAs-Al_xGa_{1-x}As superlattices," *Appl. Phys. Lett.*, vol. 47, pp. 972-974, 1985.

[26] H. Blauvelt, S. Margalit, and A. Yariv, "Single-carrier-type dominated impact ionisation in multilayer structures," *Electron Lett.*, vol. 18, pp. 375-376, 1982.

[27] J. S. Smith, L. C. Chiu, S. Margalit, A. Yariv, and A. Y. Cho, "A new infrared detector using electron emission from multiple quantum wells," *J. Vac. Sci. Technol. B*, vol. 1, pp. 376-378, 1983.

[28] S. L. Chuang and K. Hess, "Impact ionization across the conduction-band-edge discontinuity of quantum-well heterostructures," *J. Appl. Phys.*, vol. 59, pp. 2885-2894, 1986.

[29] F. Capasso, J. Allam, A. Y. Cho, K. Mohammed, R. J. Malik, A. L. Hutchinson, and D. Sivco, "New avalanche multiplication phenomenon in quantum well superlattices: Evidence of impact ionization across the band-edge discontinuity," *Appl. Phys. Lett.*, vol. 48, May 1986.

[30] F. Capasso, "The channeling avalanche photodiode A novel ultra-low-noise interdigitated p-n junction detector," *IEEE Trans. Elect. Dev.*, vol. ED-29, pp. 1388-1395, 1982.

[31] K. Brennan, "Theory of the channeling avalanche photodiode," *IEEE Trans. Elect. Dev.*, vol. ED-32, pp. 2467-2478, 1985.

[32] K. Matsuo, M. C. Teich, and B. E. A. Saleh, "Noise properties and time response of the staircase avalanche photodiode," *IEEE Trans. Electron Devices*, vol. ED-32, pp. 2615-2623, 1985. also, *J. Lightwave Technol.*, vol. LT-3, pp. 1223-1231, 1985.

[33] K. Brennan, "Theory of electron and hole impact ionization in quantum well and staircase superlattice avalanche photodiode structures," *IEEE Trans. Electron Devices*, vol. ED-32, pp. 2197-2205, 1985. "Theory of GaInAs AlInAs doped quantum-well APD A new low-noise solid-state photodetector for light-wave communications systems," *IEEE Trans. Elect. Dev.*, submitted for publication 1986.

[34] R. W. Engstrom, *RCA Photomultiplier Handbook (PMT-62)*. Lancaster, PA: RCA Electro Optics and Devices, 1980.

[35] P. M. Woodward, "A statistical theory of cascade multiplication," *Proc. Cambridge Phil. Soc.*, vol. 44, pp. 404-412, 1948.

[36] G. A. Morton and H. M. Smith, "Pulse height resolution of high gain first dynode photomultipliers," *Appl. Phys. Lett.*, vol. 13, pp. 356-357, 1968.

[37] L. Janossy, *Zh. Eksperim. Teor. Fiz.*, vol. 28, p. 679, 1955. also, "Statistical problems of an electron multiplier," *Sov. Phys.—JETP*, vol. 1, pp. 520-531, 1955.

[38] F. J. Lombard and F. Martin, "Statistics of electron multiplication," *Rev. Sci. Instrum.*, vol. 32, pp. 200-201, 1961.

[39] K. Matsuo, B. E. A. Saleh, and M. C. Teich, "Cascaded Poisson processes," *J. Math. Phys.*, vol. 23, pp. 2353-2364, 1982.

[40] J. R. Prescott, "A statistical model for photomultiplier single-electron statistics," *Nucl. Instrum. Meth.*, vol. 39, pp. 173-179, 1966.

[41] G. Lachs, "The statistics for the detection of light by nonideal photomultipliers," *IEEE J. Quantum Electron.*, vol. QE-10, pp. 590-596, 1974.

[42] C. W. Helstrom, "Output distributions of electrons in a photomultiplier," *J. Appl. Phys.*, vol. 55, pp. 2786-2792, 1984.

[43] M. Greenwood and G. U. Yule, "An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents," *J. Royal Statist. Soc.*, ser. A, vol. 83, pp. 255-279, 1920.

[44] B. E. A. Saleh, *Photoelectron Statistics*. Berlin/Heidelberg/New York: Springer-Verlag, 1978.

[45] P. R. Prucnal and M. C. Teich, "Multiplication noise in the human visual system at threshold 2 Probit estimation of parameters," *Biol. Cybern.*, vol. 43, pp. 87-96, 1982.

[46] P. R. Prucnal and B. E. A. Saleh, "Evaluation of fiber-optic error rates using a normalizing transform," *J. Opt. Soc. Amer.*, vol. 72, pp. 1171-1178, 1982.

[47] P. P. Webb and R. J. McIntyre, "Recent developments in silicon avalanche photodiodes," *RCA Eng.*, vol. 27, pp. 96-102, 1982.

[48] R. J. McIntyre, RCA Electro-Optics Photodetectors, Ste. Anne de Bellevue, P.Q., Canada, personal communication.

Malvin C. Teich (S'62-M'66-SM'72) was born in New York, NY. He received the S.B. degree in physics from the Massachusetts Institute of Technology, Cambridge, in 1961, the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 1962, and the Ph.D. degree in quantum electronics from Cornell University, Ithaca, NY, in 1966.

In 1966 he joined the M.I.T. Lincoln Laboratory, Lexington, MA, where he was engaged in work on coherent infrared detection In 1967 he became a member of the faculty in the Department of Electrical Engineering, Columbia University, New York, NY, where he is now teaching and pursuing his research interests in the areas of optical and infrared detection, quantum optics, lightwave communications, and sensory perception. He served as Chairman of the Department from 1978 to 1980. He is also a member of the faculty in the Department of Applied Physics and Nuclear Engineering, and a member of the Columbia Radiation Laboratory, the Center for Telecommunications Research, and the Columbia Bioengineering Institute. He has authored or coauthored some 100 technical publications and holds one patent

Dr. Teich is a member of Sigma Xi, the American Physical Society, the Acoustical Society of America, the Society for Neuroscience, the American Association for the Advancement of Science, and the New York Academy of Sciences. He served as a member of the Editorial Advisory Panel for *Optics Letters* from 1977 to 1979 In 1969 he was the recipient of the IEEE Browder J Thompson Memorial Prize for his paper "Infrared Heterodyne Detection" and in 1981 he received the Citation Classic Award of *Current Contents* for this work He was appointed a Fellow of the John Simon Guggenheim Memorial Foundation in 1973 and was elected a Fellow of the Optical Society of America in 1983.

Kuniaki Matsuo (M'85) was born in Hiroshima, Japan He received the B.S. degree in electrical engineering from Tokyo Electrical Engineering College, Tokyo, in 1972 and the M.S and Ph D degrees in electrical engineering from Columbia University, New York, NY, in 1978 and 1984, respectively

From 1984 to 1986 he was a Postdoctoral Research Scientist in the Columbia Radiation Laboratory, Department of Electrical Engineering, Columbia University. In 1986 he joined the Hiroshima-Denki Institute of Technology, Hiroshima, Japan, where he is now teaching and pursuing his research interests in avalanche photodetection, quantum electronics, and point processes.

Dr. Matsuo is a member of Sigma Xi.

Bahaa E. A. Saleh (M'73) received the B.S. degree from Cairo University, Cairo, Egypt, in 1966 and the Ph.D. degree from The Johns Hopkins University, Baltimore, MD, in 1971, both in electrical engineering.

From 1971 to 1974 he was an Assistant Professor at the University of Santa Catarina, Brazil. Thereafter, he joined the Max Planck Institute, Göttingen, Germany, where he was involved in research in laser light scattering and photon correlation spectroscopy. He is presently Professor of Electrical and Computer Engineering at the University of Wisconsin, Madison, where he has been since 1977. He held visiting appointments at the University of California, Berkeley, in 1977, and the Columbia Radiation Laboratory of Columbia University in 1983. He is currently involved in research in image processing, optical information processing, statistical optics, optical communication, and vision. He is the author of *Photoelectron Statistics* (Springer, 1978) and a co-editor of *Transformations in Optical Signal Processing* (SPIE, 1981) In 1980-1983 he was an Associate Editor of the *Journal of the Optical Society of America*, and since 1983 he has been a Topical Editor of the same journal

Dr. Saleh is a Fellow of the Optical Society of America and a member of Phi Beta Kappa and Sigma Xi. He was one of the recipients of the 1984-1985 Guggenheim Fellowship.

Photon-number-squeezed recombination radiation in semiconductors

Malvin C. Teich

Columbia Radiation Laboratory, Department of Electrical Engineering. Columbia University. New York. New York 10027

Federico Capasso

AT&T Bell Laboratories. Murray Hill, New Jersey 07974

Bahaa E. A. Saleh

Department of Electrical and Computer Engineering. University of Wisconsin. Madison, Wisconsin 53706

Recombination radiation from semiconductors can exhibit a squeezed photon number by use of a specially designed space-charge-limited light-emitting structure. The light will exhibit a small Fano factor and large photon flux. The structure will be small in size and have the capability of being modulated at a high rate.

We recently reported the generation of cw photon-number-squeezed (sub-Poisson) light from a Franck–Hertz experiment in which Hg vapor was excited by inelastic collisions with a space-charge-limited (quiet) electron beam.[1] The experiment operates by transferring the anticlustering properties of the electrons, resulting from Coulomb repulsion, to the photons. The direction of transfer is the inverse of that encountered in the usual photodetection process, in which the statistical character of the photons is imparted to the photoelectrons. Since the mechanism involves ordinary spontaneous emission, it is a first-order optical process and can be expected to produce a strong effect.[2]

It has been shown that the combination of sub-Poisson excitations, each associated with a single-photon emission, leads to sub-Poisson photon counts. provided that certain conditions on the characteristic times and areas of the detection process are obeyed.[2] The sub-Poisson electron excitations are characterized by a time constant τ_e that represents the time over which excitation events are anticorrelated (antibunched). The single-photon emissions, on the other hand, are characterized by a photon excitation/emission lifetime τ_p. The detected light will be sub-Poisson. provided that $T \gg \tau_e, \tau_p$ and $A \gg A_c$, where T is the detector counting time. A is the detector counting area. and A_c is the coherence area.

In spite of these conditions being satisfied in the Franck–Hertz experiment. the light was only weakly sub-Poisson, principally because of optical losses in the experimental apparatus.[1] A useful source of such light should exhibit a photon Fano factor $F_n(T)$ that is substantially less than unity and should produce a large photon flux [corresponding to a large average photon number $\langle n(T)\rangle$]. If the light is to be used in an application such as light-wave communications, the switching time (or symbol duration) T should also be able to be made small so that the device can be modulated at a high rate.[4] On the other hand. T must be sufficiently large in comparison with the characteristic times τ_e and τ_p to ensure that the sub-Poisson character of the photons is captured in the counting time.[1,2] Ideally. the device should also be small in size and produce a directed output so that the light can be coupled to an optical fiber. The structure should be designed in such a way that light loss is minimized.[3]

A number of suggestions have been made for achieving sub-Poisson excitations and single-photon emissions. Probably the simplest method is to discharge a capacitor C, containing a fixed charge. through a circuit containing a photon emitter such as a light-emitting diode (LED). The current waveform will then be a nonstationary pulse with time constant $\tau = RC$ (where R is the resistance of the circuit). A steady-state version of this experiment would make use of a fixed current stabilized by an *external* feedback circuit to drive the LED. This can be achieved by means of a constant-voltage source in series with an external resistor,[5] R. or in series with some other optoelectronic component with a suitable I-V characteristic. Other configurations. in which the anticorrelation property of the electrons is achieved by using external feedback, have been suggested.[5,6]

However, in all cases using external feedback, the characteristic anticorrelation time of the excitations τ_e is determined by the feedback time constant of the loop τ_f. A lower limit on the feedback time constant is imposed by the response time and transit time of carriers through the device and by the RC characteristics of the feedback circuitry. In general, an internal feedback process, such as space charge, will provide a more effective means of providing sub-Poisson excitations than will external feedback. This is because an internal physical process is likely to result in a smaller value of τ_e than will external electronic circuitry. Configurations

making use of space-charge-limited excitations will therefore have the capacity of being switched faster[7] than those making use of external feedback.

Using another approach altogether, it has been suggested that sub-Poisson light can be generated by the use of *nonlinear-optics* schemes that rely on correlated photon pairs and selective deletion.[8,9]   An experiment using parametric downconversion[10,11] and dead-time optical gating has indeed been successfully used to generate such light.[12]   Unfortunately this new source of sub-Poisson light suffers from a number of familiar problems: a Fano factor very close to unity, low photon flux, and slow switching speed.

It can be shown that the sub-Poisson electron-excitation methods are, in general, superior to the nonlinear-optics methods.   The advantage stems from several factors: (1) Photons naturally gravitate toward Poisson counting statistics and shot-noise fluctuations.[2,3]   It is difficult for the nonlinear-optics methods to undo this natural Poisson photon noise.   Electrons, on the other hand, are often governed by *quieter thermal-noise fluctuations,*[7,13] *thereby permitting* $F_n(T)$ to be made smaller.   (2) Nonlinear-optics schemes in which Poisson photons *are first generated (subject to a source power constraint) and subsequently converted into sub-Poisson photons cannot provide performance gain in applications such as light-wave communications.*[4] (3) *Sub-Poisson electron-excitation configurations produce light by means of efficient single-photon transitions; large values for the photon flux are therefore easily achieved.* Nonlinear-optics methods, on the other hand, rely on (relatively) inefficient multiple-photon transitions.   Furthermore, they are subject to photon interference effects that can limit the degeneracy parameter (and therefore the photon flux) to small values.[8] (4) Electron excitations, especially those mediated by a physical process such as space charge, can attain a small characteristic response time τ_e so that fast switching can be achieved.[7]   (However, as a point of interest, it should be noted that correlated photon pairs and postdetection processing, such as subtraction and correlation, may be useful in specialized applications.[14])

We conclude that an ideal source of squeezed-photon-number light will operate by means of sub-Poisson excitations, mediated by an internal physical feedback process, and will incorporate a mechanism for achieving highly efficient single-photon emissions.   It should be small in size and preferably be able to produce light in a directed beam.

We are therefore led to propose a semiconductor device structure in which sub-Poisson electron excitations are attained through space-charge-limited current flow and single-photon emissions are achieved by means of recombination radiation.   A device of this nature will emit photon-number-squeezed recombination radiation.   The energy-band diagram for such a space-charge-limited light-emitting device (SCLLED) is illustrated in Fig. 1.   Sub-Poisson electrons are directly converted into sub-Poisson photons, as in the space-charge-limited Franck–Hertz experiment,[1] but these are now recombination photons in a semiconductor. *In designing such a device, carrier and photon confinement should be optimized and optical losses should be minimized.*

The basic structure of the device is that of a p^+–i–n^+ diode.   As an example, the p^+ region may consist of $Al_{0.35}Ga_{0.65}As$, several thousand angstroms thick and doped to $p^+ \approx 5 \times 10^{17}$ cm^{-3}, and of a layer of smaller-gap material
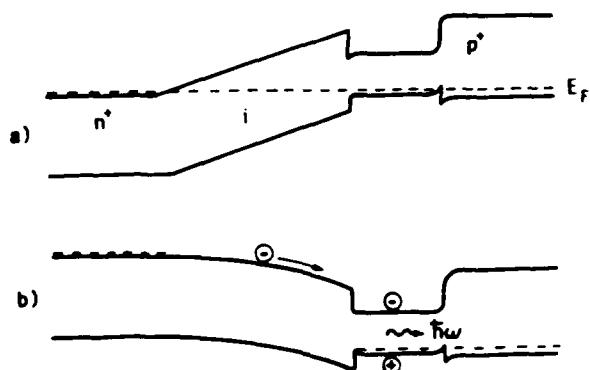

Fig. 1.   Energy-band diagram of a space-charge-limited light-emitting device under (a) equilibrium conditions and (b) strong forward-bias conditions.   The curvature of the intrinsic region under forward-bias conditions indicates the space-charge potential. E_F is the Fermi level, and $\hbar\omega$ is the photon energy.

for confinement, of GaAs with the same thickness and doping level.   The intrinsic (i) region would consist of a 2–10-μm layer of ultralow-doped (nearly intrinsic) $Al_{0.35}Ga_{0.65}As$. The n^+ region would also be of $Al_{0.35}Ga_{0.65}As$.   The structure should be grown on a conductive substrate.   A 0.15-eV discontinuity in the valence band, between the GaAs and the $Al_{0.35}Ga_{0.65}As$ layers of the p^+ region, would ensure hole confinement in the low-gap material at room temperature. The conduction-band discontinuity should be about 0.25 eV. Near-infrared recombination radiation would be emitted from the GaAs layer in this LED-like region.   This structure is appropriate for a waveguiding edge-emitting geometry. To maximize the external quantum efficiency, a surface-emitting geometry may be preferable; the surface should be dome shaped to eliminate total internal reflection.   In this case the p^+ AlGaAs layer is not necessary.

The I-V characteristic for a single-carrier conventional space-charge-limited diode is given by the approximate relationship[15]

$$\langle I_e \rangle \approx (9/8)(\epsilon_s \mu_n / L^3) A \langle V_e \rangle^2, \qquad (1)$$

where $\langle I_e \rangle$ is the average forward current, ϵ_s is the permittivity of the medium, μ_n is the electron mobility, $\langle V_e \rangle$ is the applied forward-bias voltage, and A is the device area.   This *expression is valid under strong injection conditions (applied bias voltage > built-in potential), provided that $L \gg L_D$,* where L is the thickness of the intrinsic region and L_D is the electron diffusion length.   Typical parameter values for the structure shown in Fig. 1 are $\epsilon_s \approx 1.15 \times 10^{-12}$ F/cm, $\mu_n \approx 1000$ cm^2/V sec, $L \approx 2.5$ μm, $A \approx 10^{-4}$ cm^2, $\langle V_e \rangle \approx 2$ V, and $\langle I_e \rangle \approx 33$ mA.

The current noise in such devices can be quite low.   It has a thermal (rather than shot-noise) character.[13,16]   The current-noise spectral density $S_e(\omega)$ for a device in which only electrons participate in the conduction process takes the simple form[16]

$$S_e(\omega) = 8k\theta \, d\langle I_e \rangle / d\langle V_e \rangle, \qquad (2)$$

where k is Boltzmann's constant, θ is the device temperature in degrees Kelvin, $d\langle I_e \rangle / d\langle V_e \rangle$ is the differential conductance of the device, and ω is the circular frequency.   Combining expression (1) and its derivative with Eq. (2) provides

$$S_e(\omega)/2e\langle I_e \rangle = 8k\theta/e\langle V_e \rangle, \qquad (3)$$

where e is the electronic charge. By definition, the electron current is sub-shot noise, provided that

$$S_e(\omega)/2e\langle I_e \rangle = 8k\theta/e\langle V_e \rangle < 1. \qquad (4)$$

The statistical properties of light generated by sub-Poisson electron excitations and single-photon emissions are well understood.[2] Light is considered to be sub-Poisson when the photon Fano factor $F_n(T)$ is less than unity,[17] i.e., when

$$F_n(T) = \text{var}[n(T)]/\langle n(T) \rangle < 1. \qquad (5)$$

The quantities $\text{var}[n(T)]$ and $\langle n(T) \rangle$ are the photon-number variance and photon-number mean in the counting time interval $[0, T]$, respectively. A source of light may be sub-Poisson for one value of T and super-Poisson for another value of T. In the limit of a long-counting-time/large-area detector, the photons behave as classical particles, and the expected *photoelectron* (postdetection) Fano factor $F_m(T)$ is given by[1-3]

$$[F_m(T) - 1] = \eta\beta[F_e(T) - 1], \qquad T \gg \tau_e, \tau_p; A \gg A_c. \qquad (6)$$

Here $F_e(T)$ is the Fano factor of the exciting electron stream and η is the overall quantum efficiency from electrons to detected photons. The quantity β accounts for the admixture of independent Poisson background light ($0 \leq \beta \leq 1; \beta = 1$ in the absence of such background light).[3]

The analog versions of Eqs. (5) and (6) are of interest when the detected photocurrent or the excitation current is continuous (as expected for the SCLLED) rather than when they are a sequence of discrete events. The formula analogous to Eq. (6) relates the power spectral densities of the excitation current $S_e(\omega)$ and the detected photocurrent $S_m(\omega)$. The ratios $S_j(\omega)/2e\langle I_j \rangle$ may be regarded as Fano factors $F_j(T_j)$, where $j = e, m$ and the $\langle I_j \rangle$ are the mean values of the respective currents. Here the counting times T_j play the role of inverse bandwidths of the filters involved. In the limits $T_j \gg \tau_e, \tau_p$ and for $\omega \ll 1/\tau_e, 1/\tau_p$, we obtain

$$[S_m(\omega)/2e\langle I_m \rangle - 1] = \eta\beta[S_e(\omega)/2e\langle I_e \rangle - 1],$$
$$\omega \ll 1/\tau_e, 1/\tau_p; \qquad A \gg A_c. \qquad (7)$$

It follows from Eqs. (4) and (7) that a sub-shot-noise electron-excitation current will always lead to a sub-shot-noise detected photon current. In accordance with Eq. (4), this implies sub-Poisson photon behavior when $8k\theta/e\langle V_e \rangle < 1$. For a space-charge-limited diode such as that shown in Fig. 1, it is estimated that $8k\theta/e\langle V_e \rangle \approx 0.1$ when $\theta = 300$ K and $\langle V_e \rangle = 2$ V. Further reduction of this ratio can be achieved by reducing the temperature of the device. It is interesting to note that, in the regime in which formulas (1) and (2) are valid, the degree of sub-shot-noise behavior is independent of the parameters ϵ_s, μ_n, L, and A. The values of these parameters do play a role in determining the photon flux and the device speed, however. More general results are available for the double-carrier space-charge-limited device.[15]

Using Eqs. (6) and (7), the degree of photon-number squeezing in the detected photons is then

$$F_m(T) = 1 + \eta(8k\theta/e\langle V_e \rangle - 1). \qquad (8)$$

provided that background light is absent ($\beta = 1$). The overall quantum efficiency η is given by the product $\eta_{ext}\eta_d$, where η_{ext} is the external quantum efficiency of the SCLLED (the product of the internal quantum efficiency and the geometrical collection efficiency) and η_d is the external quantum efficiency of the photodetector. Although the internal quantum efficiency of a LED is ≥ 0.5,[16] the external quantum efficiency of a dome-shaped surface emitter is typically limited to about 0.15.[19,20] If a Si p-i-n photodetector is used to detect the photon-number-squeezed recombination radiation, we can take $\eta_d \approx 0.75$, whereupon $\eta \approx 0.1125$. Finally, assuming that $8k\theta/e\langle V_e \rangle \approx 0.1$, as determined above, we arrive at an overall estimated postdetection Fano factor $F_m(T) \approx 0.899$. A commercially available LED, with $\eta_{ext} \approx 0.03$, will yield $F_m(T) \approx 0.973$. In both cases, T can be as short as ≈ 1 nsec.

The SCLLED therefore promises to provide a source of photon-number-squeezed light with properties that are significantly superior to those of the Hg-vapor space-charge-limited Franck–Hertz source[1] $[F_m(T) \approx 0.998$ with $T \approx 1$ μsec] or the dead-time-gated parametric downconversion source[12] $[F_m(T) \approx 0.9998$, with $T \approx 19$ μsec]. Indeed, the degree of photon-number squeezing of recombination radiation from the SCLLED is essentially limited only by the geometrical collection efficiency.

The waveguide geometry and superfluoresence properties (single-pass stimulated emission) of the edge-emitting LED could be advantageous for providing improved beam directionality, switching speed, spectral properties, and coupling to an optical fiber.

The question arises about whether there might be a further advantage in combining space-charge-limited current injection with a semiconductor-laser structure rather than with the LED structure considered above. This could provide increased emission efficiency and additional improvement in the parameters discussed above. This will be beneficial when the laser can be drawn into a realm of operation in which it produces a state more akin to a number state[21] than a coherent state (the coherent state has Poisson photon-number fluctuations and minimal phase fluctuations). Machita *et al.*[5,22] have shown that this mode of operation can be attained in a semiconductor-laser oscillator, within the cavity bandwidth and at high-photon-flux levels, if the pump fluctuations are suppressed below the shot-noise level. Similar suggestions have been made by Smirnov and Troshin[23] and by Carroll.[14]

In summary, we have shown that the generation of sub-Poisson light is most readily achieved by the use of sub-Poisson electron excitations, mediated by a physical mechanism such as space charge, and single-photon emissions. We have proposed a space-charge-limited light-emitting structure as a fast and compact solid-state device that operates in this manner.

ACKNOWLEDGMENTS

Squeezed States

REFERENCES

1. M. C. Teich and B. E. A. Saleh, J. Opt. Soc. Am. B **2**, 275 (1985). The initial suggestion for this experiment was put forth in M. C. Teich, B. E. A. Saleh, and D. Stoler, Opt. Commun. **46**, 244 (1983).
2. M. C. Teich, B. E. A. Saleh, and J. Peřina, J. Opt. Soc. Am. B **1**, 366 (1984); M. C. Teich and B. E. A. Saleh, Prog. Opt. (to be published).
3. M. C. Teich and B. E. A. Saleh, Opt. Lett. **7**, 365 (1982); J. Peřina, B. E. A. Saleh, and M. C. Teich, Opt. Commun. **48**, 212 (1983).
4. B. E. A. Saleh and M. C. Teich, Phys. Rev. Lett. **58**, 2656 (1987).
5. Y. Yamamoto, S. Machida, and O. Nilsson, Phys. Rev. A **34**, 4025 (1986); P. R. Tapster, J. G. Rarity, and J. S. Satchell, Europhys. Lett. **4**, 293 (1987).
6. F. Capasso and M. C. Teich, Phys. Rev. Lett. **57**, 1417 (1986).
7. E. B. Moullin, *Spontaneous Fluctuations of Voltage* (Oxford U. Press, Oxford, 1938); S. K. Srinivasan, Nuovo Cimento **38**, 979 (1965); J. R. Whinnery, Sci. Electrica **5**, 219 (1959).
8. B. E. A. Saleh and M. C. Teich, Opt. Commun. **52**, 429 (1985).
9. J. G. Walker and E. Jakeman, Proc. Soc. Photo-Opt. Instrum. Eng. **492**, 274 (1985).
10. D. C. Burnham and D. L. Weinberg, Phys. Rev. Lett. **25**, 84 (1970).
11. C. K. Hong and L. Mandel, Phys. Rev. Lett. **56**, 58 (1986).
12. J. G. Rarity, P. R. Tapster, and E. Jakeman, Opt. Commun. **62**, 201 (1987).
13. M. A. Lampert and A. Rose, Phys. Rev. **121**, 26 (1961).
14. E. Jakeman and J. G. Rarity, Opt. Commun. **59**, 219 (1986).
15. S. Sze, *Physics of Semiconductor Devices*, 1st ed. (Wiley-Interscience, New York, 1969), p. 421, Eq. (95).
16. M. A. Nicolet, H. R. Bilger, and R. J. J. Zijlstra, Phys. Status Solidi B **70**, 9 (1975); 415 (1975).
17. R. Short and L. Mandel, Phys. Rev. Lett. **51**, 384 (1983).
18. H. Kressel, M. Ettenberg, J. P. Wittke, and I. Ladany, "Laser diodes and LEDs for fiber optical communications," in *Semiconductor Devices for Optical Communication*, H. Kressel, ed., Vol. 39 of Topics in Applied Physics (Springer-Verlag, New York, 1980), Chap. 2, p. 37.
19. I. Ladany, J. Appl. Phys. **42**, 654 (1971).
20. E. G. Dierschke, L. E. Stone, and R. W. Haisty, Appl. Phys. Lett. **19**, 98 (1971).
21. P. Filipowicz, J. Javanainen, and P. Meystre, Phys. Rev. A **34**, 3077 (1986).
22. S. Machida, Y. Yamamoto, and Y. Itaya, Phys. Rev. Lett. **58**, 1000 (1987).
23. D. F. Smirnov and A. S. Troshin, Opt. Spekrosk. **59**, 3 (1985) [Opt. Spectrosc. (USSR) **59**, 1 (1985)].
24. J. E. Carroll, Opt. Acta **33**, 909 (1986).

Federico Capasso

Federico Capasso received the Doctor of Physics degree, summa cum laude, from the University of Rome (Italy) in 1973, with a thesis on stimulated Raman spectroscopy. He was a research scientist at Fondazione Bordoni from 1974 to 1976, working on the theory of nonlinear phenomena in optical fibers and on dye lasers. He joined AT&T Bell Laboratories, Murray Hill, New Jersey, in 1976, where he has since engaged in research on impact ionization, 1.3–1.6-μm detectors, liquid-phase epitaxy of III-V materials, surface passivation and defect studies, high-field transport in semiconductors, novel avalanche-photodiode structures, bipolar transistors, variable-gap heterostructures, resonant tunneling, and superlattices. He pioneered the technique of band-gap engineering and has used it extensively in the design of a new class of superlattice and heterojunction devices. In 1984, Capasso received the AT&T Bell Laboratories' Distinguished Member of Technical Staff Award and the Award of Excellence of the Society for Technical Communication. He is a fellow of the Institute of Electrical and Electronic Engineers and the American Physical Society. He has given 40 invited talks at conferences in the U.S. and abroad and is the coauthor of more than 100 papers. He has been serving on the program committees of numerous international conferences. Dr. Capasso was program cochair of the 1987 Topical Meeting on Picosecond Optoelectronics and Electronics and is chair of the Solid State Detector Technical Group of the Optical Society of America. He is a member of the Optical Society of America and the New York Academy of Sciences.

Theory of light detection in the presence of feedback

J. H. Shapiro, G. Saplakoglu, and S.-T. Ho

Department of Electrical Engineering and Computer Science. Research Laboratory of Electronics, Massachusetts Institute of Technology. Cambridge. Massachusetts 02139

P. Kumar

Department of Electrical Engineering and Computer Science. Northwestern University. Evanston. Illinois 60201

B. E. A. Saleh

Columbia Radiation Laboratory. Department of Electrical Engineering. Columbia University. New York. New York 10027

M. C. Teich

Center for Telecommunications Research. Department of Electrical Engineering. Columbia University. New York. New York 10027

The usual open-loop quantum and semiclassical theories of light detection are extended to include closed-loop operation in which there is feedback from the detector to the source. It is shown that the unmistakable signatures of nonclassical light associated with open-loop detection. such as sub-shot-noise spectra and sub-Poisson photocounts, do not carry over to closed-loop systems. This behavior is illustrated through quantitatively indistinguishable quantum and semiclassical analyses of two recent closed-loop experiments in which sub-Poisson photocount statistics were produced. It turns out that if the open-loop illumination does not require the use of quantum photodetection theory, then neither does the closed-loop illumination. Conversely. if the open-loop illumination is nonclassical. then the closed-loop behavior must be analyzed quantum mechanically. The use of nonclassical field correlations to obtain light beams that give sub-Poisson open-loop photocounts from these closed-loop arrangements is discussed and generalized into a synthesis procedure for producing light beams with arbitrary open-loop photocount statistics.

1. INTRODUCTION

The usual formulations of the quantum[1-3] and semiclassical[4-7] theories of photodetection presume open-loop configurations. i.e., that there are no feedback paths leading from the output of the photodetector to the light beam impinging on that detector. In such configurations. the qualitative and quantitative distinctions between the quantum and semiclassical theories are well understood. In the quantum theory. photocurrent and photocount randomness arise from the quantum noise in the illumination beam. whereas in the semiclassical theory the fundamental source of randomness is associated with the excitations of the atoms forming the detector. Nevertheless, the quantum theory subsumes the semiclassical theory in a natural way in that their open-loop predictions coincide exactly when the quantum field illuminating the photodetector in the former approach is in a classical state. i.e., a Glauber coherent state or a classically random mixture of such states. Inasmuch as it is only recently.[8] and with some difficulty, that light beams have been generated whose quantum statistics fall outside the classical states, it is not surprising that the semiclassical theory has continued to be the mainstay of photodetection events with classical light.

The clarity of understanding associated with open-loop photodetection does not extend to closed-loop configurations in which there is a feedback path leading from the output of the detector back to the light beam at the detector input. In this paper. we develop the quantum and semiclassical theories of light detection for closed-loop configurations.[9] The fundamental quantity of interest in these theories is the random point process formed by the photodetection event times. Thus we begin, in Section 2, with a high-level review of such processes. focusing on their application to open-loop photodetection. In particular. we introduce the doubly stochastic Poisson process (DSPP)[13-18] of semiclassical open-loop theory[5-7] and the more general self-exciting point process (SEPP)[15-18] of quantum open-loop theory. This material both establishes the analytical framework that we need for the closed-loop treatment and summarizes the semiclassical limits and nonclassical signatures of open-loop operation.

Our analysis of closed-loop photodetection begins in Section 3. Here we use the incremental point-process descriptor from Section 2 to show that both the semiclassical and quantum theories lead to SEPP's for the closed-loop case. As a result, the unmistakable open-loop signatures of nonclassical light such as sub-shot-noise spectra and sub-Pois-

son photocounts. do not carry over to closed-loop systems. We illustrate this behavior by providing quantitatively indistinguishable quantum and semiclassical explanations for two recent closed-loop experiments in which sub-Poisson photocounts were produced.[19-21] There is. nevertheless. a vital distinction between the two theories as applied to these experiments. According to the quantum theory, nonclassical open-loop beams can be extracted from these closed-loop arrangements by augmenting them to exploit the nonclassical light-beam correlations associated with photon twins[22,23] or Kerr-effect quantum nondemolition (QND) measurements.[21,24] Indeed. the former approach was recently used to produce a nonclassical open-loop beam.[25] We address these field-extraction procedures in Section 4 and show that they can be generalized to a schema for synthesizing light beams of arbitrary photocount statistics.

2. POINT PROCESSES AND OPEN-LOOP PHOTODETECTION

For the purposes of this paper. the general photodetection construct of interest. be it semiclassical or quantum, open or closed loop. takes the form shown in Fig. 1. A quasi-monochromatic paraxial scalar light beam of nominal frequency ν_0 illuminates the active region of the photodetector.[26] The quantities of interest at the detector output are the photocurrent i. and the photocount record N. The former is a train of impulses (each of area q. with q the electron charge) located at the photodetection event times $\{t_i\}$.[27] The latter counts the number of such events that have occurred in the time interval $[0, t]$ and is given by

$$N = q^{-1} \int i \, d\tau \qquad (1)$$

The event times $\{t_i\}$ that underlie both i and N comprise a random point process. For convenience. they will be arranged in increasing order with t_1 denoting the first event in the interval $[0, t]$ as sketched in Fig. 2.

A. Point-Process Statistics
Without loss of generality. we shall confine ourselves to point processes that are conditionally orderly[27]: in essence. this means that events occur one at a time. More precisely. with

$$\Delta N = N_{t+\Delta t} - N_t \qquad (2)$$

being the number of events that occur during $[t, t + \Delta t)$, a point process is conditionally orderly if its incremental probabilities obey

$$\lim_{\Delta t \to 0} \Delta t^{-1}[1 - \Pr(\Delta N = 0 | t, N_t)] = \mu_t \qquad (3)$$

$$\lim_{\Delta t \to 0} \Delta t^{-1} \Pr(\Delta N = 1 | t, N_t) = \mu_t \qquad (4)$$

$$\lim_{\Delta t \to 0} \Delta t^{-1} \Pr(\Delta N \geq 1 | t, N_t) = 0 \qquad (5)$$

function) is the conditional probability per unit time that there to be an event at t, given the history of the process up to t.[28] In general. μ_t may be an arbitrary nonnegative function of $\{t, N_t\}$. in which case Eqs. (3)-(5) constitute the incremental statistical generator of the general SEPP.[?] In its simplest form. μ_t does not depend on the event history. and Eqs. (3)-(5) then describe the Poisson process with rate function μ_t. Of interest in the sequel is the DSPP. whose incremental statistics are given by Eqs. (3)-(5) with[?]

$$\mu_t = \langle \lambda_t | t, N_t \rangle, \qquad (6)$$

where angle brackets denote expectation and λ_t is a nonnegative random process that is not directly influenced by the point process. i.e., λ_t is conditionally independent of $\{t, N_t\}$ given knowledge of $\{\lambda_\tau : \tau < t\}$. As its name implies. the DSPP is a Poisson process whose rate function is the random process λ_t. It is worth noting here that the class of SEPP's is known to be broader than that of DSPP's; this point will be illustrated below in the context of open-loop photodetection.

A useful alternative specification of point-process statistics can be made through multicoincidence rates (MCR's).[2,3,17] The kth-order MCR is

$$w_k(\tau_1, \tau_2, \ldots, \tau_k) = \lim_{\Delta t \to 0} (\Delta t)^{-k} \Pr\left(\prod_{i=1}^{k} \Delta N_{\tau_i} = 1\right) \qquad (7)$$

for $k = 1, 2, \ldots$. Basically. $w_k(\tau_1, \tau_2, \ldots, \tau_k)\Delta t^k$ is the probability that events are registered within Δt intervals about the distinct times $\tau_1, \tau_2, \ldots, \tau_k$. No ordering is implied among the $\{\tau_i\}$. nor is the possibility of additional events excluded. For a Poisson process with rate function μ. we have that

$$w_k(\tau_1, \tau_2, \ldots, \tau_k) = \prod_{i=1}^{k} \mu_{\tau_i} \qquad (8)$$

for a DSPP with random rate function λ. we have that


Fig. 1   Generic photodetection configuration.

IMPULSES OF AREA q
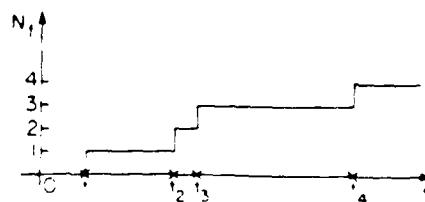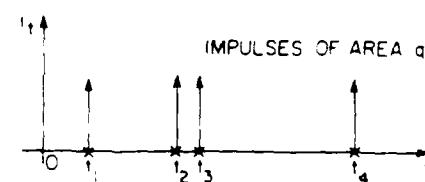
Fig. 2. Relationships between the photodetection event times, the photocurrent, and the photocount record.

$$w_k(\tau_1, \tau_2, \ldots, \tau_k) = \left\langle \prod_{i=1}^{k} \lambda_{\tau_i} \right\rangle. \tag{9}$$

There is a more general result connecting MCR's to conditional rate functions for arbitrary SEPP's, but it is not sufficiently explicit to warrant inclusion here.[17]

In terms of the MCR's, we have the following results for the photocount and photocurrent statistics associated with the point process $\{t_i\}$. The count probability distribution is

$$\Pr(N_t = n) = \sum_{m=n}^{\infty} \frac{(-1)^{m-n}}{(m-n)! \, n!} \int_0^t d\tau_1 \int_0^t d\tau_2 \ldots$$

$$\int_0^t d\tau_m w_m(\tau_1, \tau_2, \ldots, \tau_m); \tag{10}$$

it has a mean

$$\langle N_t \rangle = \int_0^t w_1(\tau) d\tau \tag{11}$$

and a variance

$$\mathrm{var}(N_t) = \langle N_t \rangle + \int_0^t d\tau \int_0^t d\tau' [w_2(\tau, \tau') - w_1(\tau) w_1(\tau')]. \tag{12}$$

For a stationary point process w_1 is a constant and w_2 depends only on the difference between its arguments. In this case the mean photocurrent is

$$\langle i \rangle = q w_1. \tag{13}$$

and its noise spectrum (bilateral covariance spectrum) is

$$S_{ii}(f) = q \langle i \rangle + q^2 \int_{-\infty}^{\infty} d\tau [w_2(\tau) - w_1^2] \exp(-i2\pi f\tau). \tag{14}$$

Equations (10)–(12) reduce to the familiar Poisson limits when the MCR's are given by Eq. (8). Their behavior for DSPP's and SEPP's forms the core of our overview of open-loop photodetection.

B. Open-Loop Photodetection

In the open-loop semiclassical theory, the event times underlying both i_t and N_t form a DSPP with random rate function given by

$$\lambda_t = \eta P_t / h\nu_0. \tag{15}$$

in terms of the detector quantum efficiency η, the (possibly random) power illuminating the detector P_t, and the photon energy $h\nu_0$. Equations (9) and (10) then yield Mandel's rule[6] for the count probability distribution

$$\Pr(N_t = n) = \left\langle \frac{W^n e^{-W}}{n!} \right\rangle, \tag{16}$$

where

$$W = \int_0^t d\tau \, \eta P_\tau / h\nu_0 \tag{17}$$

It is well known that this distribution does not permit sub-Poisson behavior, viz., from Eqs. (9), (11), (12), (15), (17),

$$\mathrm{var}(N_t) = \langle N_t \rangle + \mathrm{var}(W) \geq \langle N_t \rangle, \tag{18}$$

with equality if and only if W is nonrandom. If P_t is a stationary random process with mean $\langle P \rangle$ and noise spectrum $S_{PP}(f)$, then our DSPP formulas yield

$$\langle i \rangle = q \eta \langle P \rangle / h\nu_0 \tag{19}$$

and

$$S_{ii}(f) = q \langle i \rangle + (q\eta/h\nu_0)^2 S_{PP}(f) \tag{20}$$

for the like statistics of the photocurrent. Here we see the well-known shot-noise lower limit of semiclassical photodetection, i.e.,

$$S_{ii}(f) \geq q \langle i \rangle, \tag{21}$$

with equality at all frequencies only if P_t equals $\langle P \rangle$ with probability one.

In the open-loop quantum theory, the event times underlying i_t and N_t form a SEPP whose kth-order MCR is

$$w_k(\tau_1, \tau_2, \ldots, \tau_k) = \eta^k \, \mathrm{Tr} \left\{ \hat{\rho} \int_{A_d} d\mathbf{x}_1 \ldots \int_{A_d} d\mathbf{x}_k \right.$$

$$\left. \times \left[\prod_{i=1}^{k} \hat{E}^\dagger(\mathbf{x}_i, \tau_i) \right] \left[\prod_{i=1}^{k} \hat{E}(\mathbf{x}_i, \tau_i) \right] \right\} \tag{22}$$

where $\hat{\rho}$ is the density operator for the field, $\hat{E}(\mathbf{x}, t) \exp(-i2\pi\nu_0 t)$ is the positive-frequency photon-units field operator for $\mathbf{x} = (x, y)$ in the plane of the photodetector,[29] and A_d represents the active area of the photodetector. When $\hat{\rho}$ is a classical state, it has a proper P representation, namely,

$$\hat{\rho} = \int d^2\alpha \, P(\alpha; \alpha^*) |\alpha\rangle \langle \alpha| \tag{23}$$

for P a classical probability density and $|\alpha\rangle$ the multimode coherent state. Equation (22) then becomes

$$w_k(\tau_1, \tau_2, \ldots, \tau_k) = \int d^2\alpha P(\alpha; \alpha^*) \prod_{i=1}^{k} (\eta P_{\tau_i} / h\nu_0). \tag{24}$$

where

$$P_t = h\nu_0 \langle \alpha | \int_{A_d} d\mathbf{x} \hat{E}^\dagger(\mathbf{x}, t) \hat{E}(\mathbf{x}, t) | \alpha \rangle \tag{25}$$

is a nonnegative classical stochastic process whose statistics are specified by $P(\alpha; \alpha^*)$, and all the semiclassical formulas are reproduced. States that do not have proper P representations are called nonclassical. Their photocount and photocurrent statistics require SEPP's that, in general, are not DSPP's. For example, if $\hat{\rho}$ is a photon eigenstate that has exactly N photons within the space-time region $A_d \times [0, T)$, then Eqs. (10) and (22) yield the binomial distribution[30]

$$\Pr(N_T = n) = \binom{N}{n} \eta^n (1 - \eta)^{N-n} \tag{26}$$

for $n = 0, 1, 2, \ldots, N$. This distribution is sub-Poissonian for all $0 < \eta \leq 1$ and collapses to the expected nonrandom behavior as the quantum efficiency goes to unity.[30]

Sub-Poisson photocounts and sub-shot-noise photocurrent spectra provide unmistakable signatures for nonclassical light in open-loop photodetection. Such effects have now been observed in a variety of experiments. Short and

Mandel[10] produced conditionally sub-Poisson light by means of resonance fluorescence from sodium vapor. Their experiment required that the starting time of the photo-counting interval coincide with the entry of a single atom into the field of view of the apparatus and that the light be generated only by a single atom. Subsequently, Teich and Saleh[11] produced unconditionally (continuous-wave) sub-Poisson light by using a space-charge-limited electron beam to excite the Franck–Hertz effect in mercury vapor. More recently, Slusher *et al.*,[12] Shelby *et al.*,[13] and Maeda *et al.*[15] have seen sub-shot-noise photocurrent spectra in homodyne detection of light that has undergone four-wave mixing, as have Wu *et al.*[14] in a three-wave mixing experiment. It is relevant to note that self-exciting point processes play a role in all these cases. In the Short–Mandel experiment, the self-excitation is provided by the dead time associated with successive atomic emissions.[10] In the Teich–Saleh experiment, it is provided by the space charge associated with the electron beam. Indeed, they modeled[32] the statistics of this beam as a renewal point process, which of course, is a special kind of SEPP. Self-excitation, in the multiwave mixing experiments, can be attributed to the quantum correlations between the pump and probe beams produced by photon exchanges occurring through the nonlinear optical process.

3. CLOSED-LOOP PHOTODETECTION

Most theoretical results for open-loop systems do not carry over to closed-loop photodetection. We proceed now to reexamine the semiclassical and quantum photodetection properties of i_t and N_t when the output of the photodetector is permitted to affect the light at its input through a causal, but possibly nonlinear, feedback loop, as shown in Fig. 3. In both the semiclassical and quantum formulations it then turns out that the photodetection event times form a SEPP. Semiclassically, this is most easily seen through the incremental point-process description [Eqs. (3)–(5)]. In the ab-sence of feedback we have a DSPP, i.e., the semiclassical conditional rate obeys

$$\mu_t = \langle \eta P_t / h\nu_0 | t_-, N_{t-} \rangle \qquad (27)$$

where P_t, the illumination power, is a nonnegative random process that is not directly influenced by the point process, viz., P_t is conditionally independent of the photocount history $\{t_-, N_{t-}\}$ given its own history $\{P_\tau : \tau < t\}$. Closing the feed-back loop leaves Eq. (27) formally intact but makes

$$P_t = P_t(t_-, N_{t-}) \qquad (28)$$

an explicit function of the photocount history. Thus the DSPP condition is violated, indicating that closed-loop semiclassical photodetection leads, in general, to a SEPP.

Quantum mechanically, open-loop photodetection al-

ready requires SEPP descriptions; passing to the closed-loop configuration modifies many of the detailed results but does not break out of the general SEPP structure. For example, in the open-loop configuration it is known[4,5] that the photocurrent i_t realizes the quantum measurement[46]

$$\hat{i}_t \equiv q \int_{A_s} d\mathbf{x} \hat{E}'^\dagger(\mathbf{x}, t) \hat{E}'(\mathbf{x}, t), \qquad (29)$$

with

$$\hat{E}'(\mathbf{x}, t) \equiv \eta^{1/2} \hat{E}(\mathbf{x}, t) + (1 - \eta)^{1/2} \hat{E}_v(\mathbf{x}, t), \qquad (30)$$

where $\hat{E}_v(\mathbf{x}, t)$ is a vacuum-state field operator representing subunity quantum efficiency noise. In Eqs. (29) and (30), $\hat{E}(\mathbf{x}, t)$ and $\hat{E}_v(\mathbf{x}, t)$ are commuting free-field operators that do not explicitly depend on $\{i_\tau : \tau < t\}$. When the loop is closed, Eqs. (29) and (30) continue to apply but with $\hat{E}(\mathbf{x}, t)$ explicitly dependent on past measurements.

Because both the semiclassical and quantum theories lead to SEPP's, there are no unmistakable signatures for nonclas-sical light in the closed loop. To be specific, for a SEPP the quantity $w_2(t, s) - w_1(t)w_1(s)$ need not be positive semide-finite, so that [see Eqs. (12) and (14)] sub-Poisson photo-counts and sub-shot-noise photocurrent spectra are possible in both theories. In what follows, we shall explore this behavior for two feedback configurations, which we term the dead-time-modified Poisson process (DTMPP) and the neg-ative-linear-feedback process (NLFP).

A. Dead-Time-Modified Poisson Process

It has long been known that semiclassical photocounting statistics are impacted by dead time in the detection system. Consider the nonparalyzable DTMPP, which is a renewal process and hence a SEPP.[17,34–36] For a detector with fixed dead time τ_d, illuminated, in the semiclassical theory, by nonrandom light of constant power P, the photocount prob-ability distribution is[35,36]

$$\Pr(N_t = n) = \begin{cases} \sum_{k=0}^{n} p_0[k, \lambda(t - n\tau_d)] - \sum_{k=0}^{n-1} p_0\{k, \lambda[t - (n - 1)\tau_d]\} & \text{for } n < t/\tau_d \\[2mm] 1 - \sum_{k=0}^{n-1} p_0\{k, \lambda[t - (n - 1)\tau_d]\} & \text{for } t/\tau_d \leq n < t/\tau_d + 1, \quad (31) \\[2mm] 0 & \text{for } n \geq t/\tau_d + 1 \end{cases}$$

where

$$p_0(k, a) \equiv a^k e^{-a}/k!, \qquad \lambda \equiv \eta P/h\nu_0. \qquad (32)$$

These results are valid for a detector that is unblocked at the beginning of each counting interval, although exact number distributions for counters that are blocked and for equilibri-


INPUT LIGHT → OPTICAL MODULATOR → IN-LOOP LIGHT → PHOTO-DETECTOR → PHOTOCURRENT

CONTROL SIGNAL

CAUSAL SYSTEM
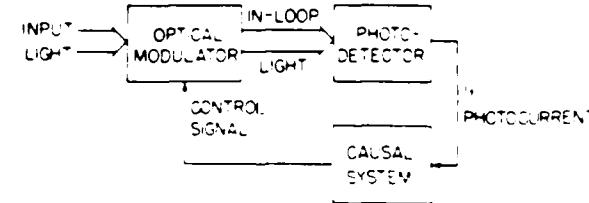
Fig. 3. Closed-loop photodetection configuration.

CAN THE CHANNEL CAPACITY OF A LIGHT-WAVE COMMUNICATION SYSTEM BE INCREASED BY THE USE OF PHOTON-NUMBER–SQUEEZED LIGHT?

B. E. A. Saleh and M. C. Teich

Can the Channel Capacity of a Light-Wave Communication System Be Increased by the Use of Photon-Number–Squeezed Light?

B. E. A. Saleh

*Department of Electrical and Computer Engineering, University of Wisconsin,
Madison, Wisconsin 53706*

and

M. C. Teich

*Columbia Radiation Laboratory and Center for Telecommunications Research,
Department of Electrical Engineering, Columbia University, New York, New York 10027*

(Received 9 January 1987)

The channel capacity C of a light-wave communication system based on photoevent point-process observations *cannot* be increased by use of photon-number–squeezed light. Under a constraint of maximum photon rate λ_{max}, the channel capacity $C = \lambda_{max}/e$ is achieved with a Poisson process. On the other hand, the channel capacity of a communication system based on photon counting *can* be increased by use of photon-number–squeezed light. The improvement vanishes in the limit of very small mean counts.

PACS numbers 42.50.Dv

All light-wave communication systems that have been developed to date make use of Poisson (or super-Poisson) light.[1] For Poisson light, the variance of the photon number is identically equal to its mean for all values of the counting time T. Photon-number–squeezed light, on the other hand, has a photon-number variance that is less than its mean for all or some values of T.[2,3] Such light is intrinsically nonclassical in nature. The earliest source of *unconditionally* photon-number–squeezed (also called sub-Poisson) light exhibited only a slight reduction of the variance.[4] Stronger photon-number squeezing has been produced in more recent experiments[5,6] and continuing developments[7-9] promise further improvement in the future. It is therefore of interest to investigate whether there might be an advantage to using photon-number–squeezed light in a direct-detection light-wave communication system.

There are two classes of mechanisms by means of which unconditionally photon-number–squeezed light may be generated. In the first class, squeezed photons are produced from a beam of *initially Poisson* (or super-Poisson) *photons*. This can be achieved in a number of ways, e.g., by the use of correlated photon beams or quantum nondemolition (QND) measurements.[10] An experiment of this kind was recently carried out by Rarity, Tapster, and Jakemen.[11] Sub-Poisson photons were generated from the pair of correlated photon beams produced in parametric down-conversion; one of the twin beams was used to gate the other beam selectively via dead-time control.

The second class of mechanisms relies on the direct generation of squeezed photons from a beam of *initially sub-Poisson excitations* (e.g., electrons).[3,12] This technique was first used by Teich and Saleh in a space-charge-limited version of the Franck-Hertz experiment.[4]

Perhaps the simplest implementation of this principle is achieved by the driving of a light-emitting diode with a sub-Poisson electron current.[6,8]

In this Letter we show that the channel capacity of a light-wave communication system based on the observation of the *photoevent point process cannot* in principle be increased by the use of photon-number–squeezed light. On the other hand, the channel capacity of a *photon-counting* system *can* be increased by the use of photon-number–squeezed light. The channel capacity is the maximum rate of information that can be transmitted through a channel without error.[13] We also provide an example in which the use of photon-number–squeezed light produced from Poisson light either degrades or enhances the *error performance* of a simple binary on-off keying photon-counting system, depending on where the average power constraint is placed.

Consider the transformation of a Poisson beam of photons (represented by a Poisson point process[10] N_t of rate μ_t) into a sub-Poisson beam of photons represented by a point process M_t of rate λ_t. The events of the initial process N_t are assumed to be observable (e.g., by the use of correlated photon beams or a QND measurement) and their registrations used to operate a mechanism which, in accordance with a specified rule, leads to the events of the transformed photon process M_t. The rate λ_t of the process M_t is thereby rendered a function of the realizations of the initial point process N_t at prior times, i.e., $\lambda_t = \lambda_t(N_{t'}; t' \leq t)$.

Several examples of transformations of this kind that have been suggested for use in quantum optics are illustrated in Fig. 1 and discussed below. It is assumed for simplicity (but without loss of generality) that the various conversions can be achieved in an ideal manner.

(i) *Dead-time deletion.* Delete all photons within a
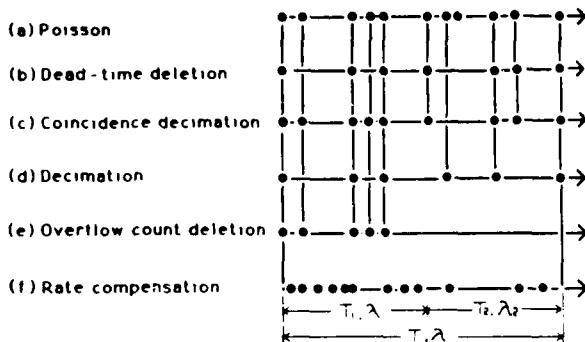
FIG. 1. Several transformations of Poisson photons into sub-Poisson photons that have been suggested for use in quantum optics.

prescribed fixed (nonparalyzable) dead time τ following the registration of a photon.[14] Rarity, Tapster, and Jakeman[11] generated photon-number–squeezed light by using one of the twin beams produced in parametric down-conversion to gate photons selectively from the other beam via dead-time control. Dead-time deletion could also be used with correlated photon beams produced in other ways.

(ii) *Coincidence decimation:* Remove all pairs of photons separated by a time shorter than a prescribed time interval τ'. This is achieved, for example, in second-harmonic generation; two photons closer than the intermediate-state lifetime of the second-harmonic-generation process are exchanged for a third photon (which is at twice the frequency and therefore easily eliminated).[15]

(iii) *Decimation:* Select every rth photon ($r = 2$, 3, . . .) of an initially Poisson photon process, deleting all intermediate photons. Saleh and Teich[16] suggested using correlated photon beams to implement this technique. In cascaded atomic emissions from ^{40}Ca, for example, sequences of correlated photon pairs (green and violet) are emitted. The green photons can be detected and used to operate a gate that passes every rth violet photon. Decimation control could also be used in conjunction with parametric–down-conversion photon twins.

(iv) *Overflow count deletion:* The number of photons occurring in preselected time intervals $[0, T_0]$, $[T_0, 2T_0]$, . . . , is counted, retaining the first n_0 photons in each time interval (without changing their occurrence times) and deleting the remainder. If the average number of photons in $[0, T_0]$ of the initial process is $\gg n_0$, then the transformed process will almost always contain n_0 photons within this time interval. As an example, Mandel[17] suggested that if a collection of n_0 atoms in the ground state are subjected to a brief, intense, incoherent excitation pulse, all n_0 atoms will become excited with high probability; the radiated optical field would

then be describable, to good approximation, by an n_0-photon state. Related schemes have been proposed by Yuen[18] and by Stoler and Yurke[19] for use with parametric processes.

We proceed to illustrate that none of these modifications can increase the channel capacity of a communication system based on photoevent point-process observations.

If a constraint is placed on the rate of the initial Poisson process $\mu_t \leq \mu_{max}$, then it is obvious that C cannot be increased by the modification $N_t \to M_t$. This is simply a consequence of the definition of channel capacity: It is the rate of information carried by the system without error, maximized over all coding, modulation, and *modification* schemes. Can the modification $N_t \to M_t$ increase the channel capacity if the constraint is instead placed on the rate of the modified process λ_t (i.e., $\lambda_t \leq \lambda_{max}$)?

We address this question for an arbitrary self-exciting point process M_t of rate $\lambda_t(M_{t'}; t' \leq t)$. This is a process that contains an inherent feedback mechanism in which present event occurrences are affected by the previous event occurrences of the same point process. Of course, the modified Poisson processes $N_t \to M_t$ introduced above are special cases of self-exciting point processes.

An example of a system that generates a self-exciting point process is that of rate compensation (by linear feedback) of a source which, without feedback, would produce a Poisson process. Let each photon registration at time t_i cause the rate of the process to be modulated by a factor $h(t - t_i)$ (which vanishes for $t < t_i$). In linear negative feedback the rate is $\lambda_t = \lambda_0 - \sum_i h(t - t_i)$, where λ_0 is a constant. If the instantaneous photon registration rate happens to be above the average then it is reduced, and vice versa. This process is schematically illustrated in Fig. 1(f) for two adjacent subintervals T_1 and T_2. Yamamoto, Imoto, and Machida[20] suggested the use of rate compensation in conjunction with a QND measurement (using the optical Kerr effect), but it could be used just as well, for example, with correlated photon pairs. Dead-time deletion can be viewed as a special case of rate compensation in which the occurrence of an event zeros the rate of the process for a specified time period τ after the registration.[10]

Now consider a communication system that uses a point process $M_t(X)$ whose rate $\lambda_t(X)$ is modulated by a signal X_t. The process $M_t(X)$ can be an arbitrary self-exciting point process (e.g., it can be photon-number squeezed) which includes processes obtained by the feedforward or feedback modification of an otherwise Poisson process. Neither feedforward nor feedback transformations can increase the capacity of this channel, as provided by Kabanov's theorem.[21]

Kabanov's theorem.—The capacity of the point-process channel cannot be increased by feedback. Under

the constraint $\lambda_l \leq \lambda_{max}$, the channel capacity C is

$$C = \lambda_{max}/e. \tag{1}$$

When the capacity is achieved, the output of the point-process channel is a Poisson process with rate $\lambda_l = \lambda_{max}/e$ (the base e has been used for simplicity). The channel capacity has also been determined under added constraints on the minimum rate (dark events) and on the mean rate.[21] A coding theorem has also been proved.

In summary, *no* increase in the channel capacity of a point-process light-wave communication system may be achieved by the use of photons that are first generated with Poisson statistics and subsequently converted into sub-Poisson statistics regardless of whether the power constraint is placed at the Poisson photon source or at the output of the conversion process. Nor may an increase in channel capacity be achieved by the use of feedback to generate a self-exciting point process.

These conclusions are valid only when there are no restrictions on the receiver structure. The conclusion is different if the receiver is operated in the *photon-counting* regime, in which information is carried by a random variable n representing the number of photo-events registered in time intervals of prescribed duration T (rather than by the photon occurrence times).

The capacity of the photon-counting channel is given by[13]

$$C = B[\bar{n}\ln(1 + 1/\bar{n}) + \ln(1 + \bar{n})]. \tag{2}$$

where $\bar{n}$ is the mean number of counts in T and $B = 1/T$ is the bandwidth. Two limiting expressions emerge:

$$C = \begin{cases} B\bar{n}\ln(1/\bar{n}), & \bar{n} \ll 1, \\ B\ln(\bar{n}), & \bar{n} \gg 1. \end{cases} \tag{3}$$

If an added constraint is applied to the photon counts, such that they must obey the Poisson counting distribution, the capacity is further reduced. In that case, the limiting results analogous to Eq. (3) are

$$C = \begin{cases} B\bar{n}\ln(1/\bar{n}), & \bar{n} \ll 1, \\ \frac{1}{2}B\ln(\bar{n}), & \bar{n} \gg 1. \end{cases} \tag{4}$$

In the case of photon counting, therefore, an increase in the channel capacity *can* in principle be realized by the use of photon-number–squeezed light. However, in the small mean-count limit $\bar{n} \ll 1$ (very short T), the capacity of the Poisson counting channel approaches that of the unrestricted counting channel, and the advantage of photon-number squeezing disappears. This is not unexpected in view of the result obtained from Kabanov's theorem for the point-process channel.

The channel capacity provides a limit on the maximum rate of error-free information transmission for all codes, modulation formats, and receiver structures.[13] As such, it does not specify the performance (error probability) achievable by a communication system with pre-scribed coding, modulation, and receiver structure.

It is therefore of interest to examine the performance of a system with specified structure. We consider a binary on-off keying photon-counting system.[1] The information is transmitted by the selection of one of two values for the photon rate λ_l, in time slots of (binary-digit) duration T. The receiver operates by counting the number of photons received during the time interval T and then deciding which rate was transmitted in accordance with a likelihood-ratio decision rule (threshold test). For simplicity, it is assumed that background light, dark noise, and thermal noise are absent so that photon registrations are not permitted when the keying is OFF (i.e., false alarms are not possible). Furthermore, the detector quantum efficiency is taken to be unity so that system performance is limited only by the quantum fluctuations of the light.

A measure of performance for a digital system such as this is the error probability P_e. In the simplified system described above, errors are possibly only when the keying is ON and 0 photons are received (a miss). For a Poisson transmitter, P_e is[1]

$$P_e(\text{Poisson}) = \frac{1}{2}\exp(-\bar{n}). \tag{5}$$

where $\bar{n}$ denotes the mean number of emitted photons. To minimize P_e, $\bar{n}$ is made equal to its maximum allowed value $\bar{n}_{max}$. This result is now compared with those obtained for photon-number–squeezed light derived from an initially Poisson source. The outcome will depend on where the mean photon-number constraint is placed. Two transformations are explicitly considered: dead-time deletion and decimation.

(i) *Dead-time deletion:* For a nonparalyzable dead-time modifier that is always *blocked* for a dead-time period τ at the beginning of the counting interval T, the passage of 0 photons arises from the emission of 0 photons in the time $T - \tau$, independent of the number of emissions during τ. The error probability for this system is therefore

$$P_e(\text{dead time}) = \frac{1}{2}\exp[-\bar{n}(1 - \tau/T)]. \tag{6}$$

To minimize error under the constraint $\bar{n} \leq \bar{n}_{max}$, we take $\bar{n} = \bar{n}_{max}$. The error is obviously larger than that for the Poisson channel [Eq. (5)] and so no performance enhancement can be achieved by use of this modifier. If, instead, the dead-time modifier is always *unblocked* at the beginning of each bit then the passage of 0 photons arises from the emission of 0 photons in the time T, and the dead time has no effect on the error rate in this simple system. Calculations for the unblocked counter in the presence of false alarms, however, demonstrate that the presence of deal time always does, in fact, degrade system performance with such a constraint.[22] Although the detailed calculations were carried out for electrical dead time, the results are also applicable for optical dead time when the photon detection efficiency $\eta = 1$. On the

other hand, if the constraint is placed on the mean photon count $\bar{m}$ after dead-time modification ($\bar{m} \leq \bar{m}_{max}$), it can be shown that there exists a value of $\bar{m}_{max}$ below which performance is degraded and above which performance is improved, relative to the Poisson channel.

(ii) *Decimation*: We assume that the decimation parameter $r = 2$ (i.e., every other photon of a Poisson sequence of events is selected) and that the decimation process is reset at the beginning of each bit (i.e., the first photon in each bit is not selected). The error probability is then

$$P_e(\text{decimation}) = \tfrac{1}{2}(1+\bar{n})\exp(-\bar{n}), \tag{7}$$

which again represents a degradation of performance in comparison with the Poisson channel (under a constraint $\bar{n} \leq \bar{n}_{max}$). In this case, the error rate is increased because there are two ways for the passage of 0 photons to arise in the time T from the emission of 0 photons or from the emission of 1 photon. However, if the constraint is placed on the modified process, then, once again, there exists a value of $\bar{m}_{max}$ below which performance is degraded and above which it is improved, relative to the Poisson channel.

It is evident from these examples that system performance can be enhanced by the use of photon-number–squeezed light, provided that the average power constraint is applied to the squeezed light. No enhancement of system performance emerges in converting Poisson photons into squeezed photons when the average power constraint is at the Poisson source.

Losses have been ignored in the considerations above. It is important to keep in mind the inevitable random photon deletion that results from absorption, scattering, and the finite quantum efficiency of the detector. It is well known that these deletions will transform a deterministic photon number into a binomial photon-number distribution,[17,23] which always remains sub-Poisson but approaches the Poisson boundary as the random deletion increases.[23] Mandel[17] has shown that the information rate per symbol carried by such a counting channel will be greater than that for the Poisson channel, but will approach the latter as the photon-detection probability η approaches 0. We have shown elsewhere[24] that the performance of a simple binary on-off–keying photon-counting system, of the kind considered earlier, is also superior for a binomial source than for a Poisson source, approaching the latter as η approaches 0.

[1] P. S. Henry, IEEE J. Quantum Electron. 21, 1862 (1985).

[2] R. Short and L. Mandel, Phys. Rev. Lett. 51, 384 (1983).

[3] M. C. Teich, B. E. A. Saleh, and D. Stoler, Opt. Commun. 46, 244 (1983).

[4] M. C. Teich and B. E. A. Saleh, J. Opt. Soc. Am. B 2, 275 (1985).

[5] S. Machida, Y. Yamamoto, and Y. Itaya, Phys. Rev. Lett. 58, 1000 (1987).

[6] P. R. Tapster, J. G. Rarity, and J. S. Satchell, Europhys. Lett. (to be published).

[7] F. Capasso and M. C. Teich, Phys. Rev. Lett. 57, 1417 (1986).

[8] M. C. Teich, F. Capasso, and B. E. A. Saleh, J. Opt. Soc. Am. B 4 (to be published) (special issue on squeezed states).

[9] Y. Yamamoto, S. Machida, and O. Nilsson, Phys. Rev. A 34, 4025 (1986).

[10] J. H. Shapiro, G. Saplakoglu, S.-T. Ho, P. Kumar, B. E. A. Saleh, and M. C. Teich, in Ref. 8.

[11] J. G. Rarity, P. R. Tapster, and E. Jakeman, Europhys. Lett. (to be published).

[12] M. C. Teich, B. E. A. Saleh, and J. Peřina, J. Opt. Soc. Am. B 1, 366 (1984).

[13] T. E. Stern, IRE Trans. Inf. Theory 6, 435 (1960); J. P. Gordon, Proc. IRE 50, 1898 (1962); J. R. Pierce, E. C. Posner, and E. R. Rodemich, IEEE Trans. Inf. Theory 27, 61 (1981); Y. Yamamoto and H. A. Haus, Rev. Mod. Phys. 58, 1001 (1986).

[14] M. C. Teich and G. Vannucci, J. Opt. Soc. Am. 68, 1338 (1978).

[15] R. Loudon, *The Quantum Theory of Light* (Clarendon, Oxford, 1983), 2nd ed.

[16] B. E. A. Saleh and M. C. Teich, Opt. Commun. 52, 429 (1985).

[17] L. Mandel, J. Opt. Soc. Am. 66, 968 (1976).

[18] H. P. Yuen, Phys. Rev. Lett. 56, 2176 (1986).

[19] D. Stoler and B. Yurke, Phys. Rev. A 34, 3143 (1986).

[20] Y. Yamamoto, N. Imoto, and S. Machida, Phys. Rev. A 33, 3243 (1986).

[21] Yu. M. Kabanov, Theory Probab. Its Appl. (Engl. Transl.) 23, 143 (1978); M. H. A. Davis, IEEE Trans. Inf. Theory 26, 710 (1980); A. A. Lazar, in Proceedings of the Fourteenth Annual Conference on Information Sciences and Systems, Princeton, NJ, March 26-28, 1980 (unpublished); P. Brémaud, unpublished.

[22] M. C. Teich and B. I. Cantor, IEEE J. Quantum Electron. 14, 993 (1978).

[23] M. C. Teich and B. E. A. Saleh, Opt. Lett. 7, 365 (1982).

[24] M. C. Teich and B. E. A. Saleh, Prog. Opt. 25 (to be published).

[36]W. Viehmann and A. G. Eubanks. NASA Report No. TN D-8147, 1976

[37]W. Viehmann, A. G. Eubanks, G. F. Piper, and J. H. Bredekamp. Appl. Opt. **14**, 2104, 1975.

[38]P. R. Tapster, J. G. Rarity, and J. S. Satchell, Europhys. Lett. **4**, 293 (1987).

[39]S. Machida, Y. Yamamoto, and Y. Itaya, Phys. Rev. Lett. **58**, 1000 (1987).

[40]L. A. Wu, H. J. Kimble, J. L. Hall, and H. Wu, Phys. Rev. Lett. **57**, 2520 (1986).

$$V(x) = \frac{\partial^2}{\partial s^2} K_x(s) \Big|_{x=0} , \tag{B4b}$$

$$M_3(x) = -\frac{\partial^3}{\partial s^3} K_x(s) \Big|_{s=0} . \tag{B4c}$$

This simple pattern does not persist beyond third order, however.

Consider a primary process p characterized by the MGF $Q_p(s)$, whose members independently initiate a branching process resulting in an overall multiplication m described by $Q_m(s)$. The moment-generating function of the overall compound process mixes the above via

$$Q_c(s) = Q_p(-\ln Q_m(s)) , \tag{B5}$$

which, using Eq. (B3), can be rewritten in terms of the associated CGF's as

$$K_c(s) = K_p(-K_m(s)) . \tag{B6}$$

Differentiating both sides according to Eq. (B4), and expanding via the chain rule, leads to the general relations

$$\langle n_c \rangle = \langle p \rangle \langle m \rangle , \tag{B7a}$$

$$V(n_c) = V(p)\langle m \rangle^2 + \langle p \rangle V(m) , \tag{B7b}$$

$$M_3(n_c) = M_3(p)\langle m \rangle^3 + 3V(p)\langle m \rangle V(m) + \langle p \rangle M_3(m) , \tag{B7c}$$

which may be applied to any simple compound process where a primary counting distribution $P(p)$ drives a multiplication counting distribution $P(m)$.

[1] B. B. Rossi, *Cosmic Rays* (McGraw-Hill, New York, 1964).

[2] Traces of protons, deuterons, tritons, alpha particles, and heavier nuclei have also been detected at ground level, as discussed by P. C. M. Yock, Phys. Rev. D **34**, 698 (1986).

[3] G. R. Smith, M. Ogmen, E. Buller, and S. Standil, Phys. Rev. Lett. **50**, 2110 (1983).

[4] H. J. Bhabha and W. Heitler, Proc. R. Soc. London **A159**, 432 (1937).

[5] J. F. Carlson and J. R. Oppenheimer, Phys. Rev. **51**, 220 (1937).

[6] B. E. A. Saleh, *Photoelectron Statistics* (Springer, New York, 1978).

[7] D. L. Snyder, *Random Point Processes* (Wiley, New York, 1975).

[8] U. Fano, Phys. Rev. **72**, 26 (1947).

[9] W. Furry, Phys. Rev. **52**, 569 (1937).

[10] G. U. Yule, Phil. Trans. R. Soc. London **B213**, 21 (1924).

[11] N. Arley, *On the Theory of Stochastic Processes and Their Application to the Theory of Cosmic Radiation* (Wiley, New York, 1943).

[12] S. K. Srinivasan, *Stochastic Theory and Cascade Processes* (Elsevier, New York, 1969).

[13] K. Shimoda, H. Takahasi, and C. Townes, J. Phys. Soc. Jpn. **12**, 686 (1957).

[14] E. Parzen, *Stochastic Processes* (Holden-Day, San Francisco, 1962).

[15] W. Shockley and J. R. Pierce, Proc. IRE **26**, 321 (1938).

[16] R. E. Burgess, J. Phys. Chem. Solids **22**, 371 (1961).

[17] C. C. Shih, Phys. Rev. D **34**, 2710 (1986).

[18] K. Matsuo, M. C. Teich, and B. E. A. Saleh, J. Math. Phys. **25**, 2174 (1984).

[19] M. C. Teich, K. Matsuo, and B. E. A. Saleh, IEEE Trans. Electron Devices **ED-33**, 1475 (1986).

[20] M. H. Quenouille, Biometrics **5**, 162 (1949).

[21] M. S. Bartlett, *An Introduction to Stochastic Processes* (Cambridge University Press, Cambridge, England, 1966).

[22] J. V. Jelley, *Čerenkov Radiation and its Applications* (Pergamon, New York, 1958).

[23] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975).

[24] A. T. Young, Appl. Opt. **8**, 2431 (1969).

[25] A. G. Wright, J. Phys. E **16**, 300 (1983).

[26] R. L. Jerde, L. E. Peterson, and W. Stein, Rev. Sci. Instrum. **38**, 1387 (1967).

[27] A. T. Young, Rev. Sci. Instrum. **37**, 1472 (1966).

[28] W. G. Fastie, Apollo 17 UV Spectrometer Experiment (S-169), final report, Addendum C, NASA Contract No. NAS 9-11528, 1974.

[29] R. W. Engstrom, *RCA Photomultiplier Handbook (PMT-62)* (RCA Electro-Optics and Devices, Lancaster, PA, 1980).

[30] It is well known that a random variable formed by the addition of N independent random variables has a moment-generating function given by the product of the individual moment-generating functions (Ref. 14).

$$Q_s(s) = \prod_{j=1}^{N} Q_j(s) .$$

This can be rewritten as

$$Q_s(s) = \exp\left[\sum_{j=1}^{N} \ln Q_j(s)\right] ,$$

which, with the help of Eq. (B3), becomes

$$K_s(s) = \sum_{j=1}^{N} K_j(s) .$$

It follows immediately from the differentiation properties of the cumulant-generating function provided in Eq. (B4) that the means, variances, and third central moments are additive, i.e.,

$$M_i(x) = \sum_{j=1}^{N} M_i(x_j) .$$

[31] J. T. Tavolacci and M. C. Teich (unpublished).

[32] This reflects the fact that each model is governed by three free parameters, so at least three independent data points per counting distribution are required for a fit.

[33] M. C. Teich and B. E. A. Saleh, J. Opt. Soc. Am. B **2**, 275 (1985).

[34] B. E. A. Saleh, J. T. Tavolacci, and M. C. Teich, IEEE J. Quantum Electron. **QE-17**, 2341 (1981).

[35] L. W. Howell and H. F. Kennel, NASA Technical Report No. 2337, 1984.

(a)

(b)
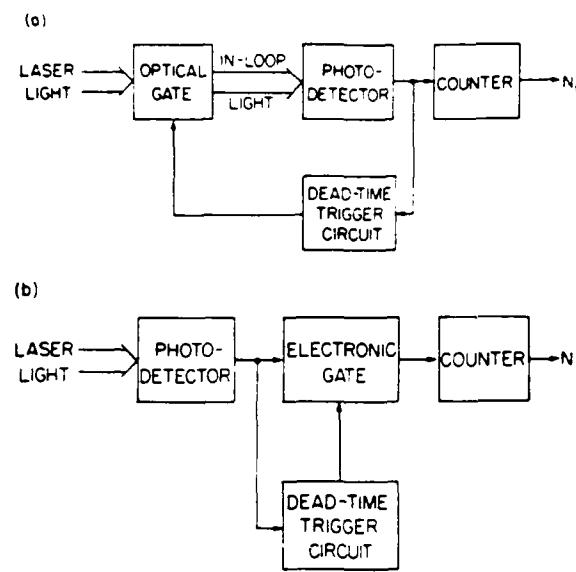
Fig. 4. (a) Closed-loop photocounting experiment carried out by Walker and Jakeman.[19] (b) Photocounting experiment carried out by Teich and Vannucci.[16]

um counters are also available. In the usual situation, the mean count is much greater than unity, in which case the differences arising from the three initial conditions are insubstantial and a simple approximation for the photocount distribution suffices.[16] The photocount mean and variance then take the asymptotic forms

$$\langle N \rangle = \lambda t/(1 + \lambda \tau_a) \tag{33}$$

and

$$\mathrm{var}(N) = \langle N \rangle/(1 + \lambda \tau_a)^2. \tag{34}$$

representing sub-Poisson behavior for all values of $\lambda \tau_a$.

The DTMPP results are relevant to experiments recently carried out by Walker and Jakeman.[19] The simplest form of their experimental arrangement is illustrated in Fig. 4(a). The registration of a photoevent at the detector operates a trigger circuit that causes an optical gate to be closed for a fixed period of time τ_d after the time of registration. During this period, the power P_i of the (He–Ne) laser illuminating the detector is set precisely equal to zero so that no photodetections are registered. The arrangement is therefore equivalent to the one illustrated in Fig. 4(b), in which the gating is electronic rather than optical, at least so far as the photocount statistics are concerned. This latter arrangement was used by Teich and Vannucci,[16] and sub-Poisson photocounts were observed in both cases. This is because the point process seen by the counter in these experiments is the DTMPP considered above. As Walker and Jakeman understood, their observations can be explained without recourse to the quantum theory of photodetection; under closed-loop conditions, sub-Poisson photocounts are possible within the semiclassical framework. Nevertheless, it is instructive to develop the quantum foundations for Eqs. (33) and (34) in a manner conducive to generalization.

Consider the quantum version of a somewhat richer dead-

time experiment shown in Fig. 5. Here, a coherent-state signal field $\hat{E}_S(\mathbf{x}, t)$ with mean

$$\langle \alpha | \hat{E}_S(\mathbf{x}, t) | \alpha \rangle = (P/h\nu_0 A_d)^{1/2} \tag{35}$$

illuminates an in-loop photodetector of quantum efficiency η through a feedback-controlled flip mirror. For τ_d sec after each in-loop photocount registration, the flip mirror directs $\hat{E}_S(\mathbf{x}, t)$ to an out-of-loop photodetector of matched quantum efficiency η. During this dead-time interval, the in-loop detector is illuminated by a vacuum-state field operator $\hat{E}_M(\mathbf{x}, t)$. After this dead-time interval the mirror returns to its previous position, in which $\hat{E}_S$ illuminates the in-loop detector and $\hat{E}_M$ illuminates the out-of-loop detector.

We derive the steady-state first and second moments (means, variances, and covariance) of the in-loop and out-of-loop photocount records N_t and N_t' in the high-mean-count limit. Toward that end, we recast the flip-mirror subsystem into the explicit lossless modulated beam-splitter form, shown in Fig. 6, in which

$$\hat{E}_{\mathrm{IN}}(\mathbf{x}, t) = [T(t)]^{1/2}\hat{E}_S(\mathbf{x}, t) + [1 - T(t)]^{1/2}\hat{E}_M(\mathbf{x}, t) \tag{36}$$

and
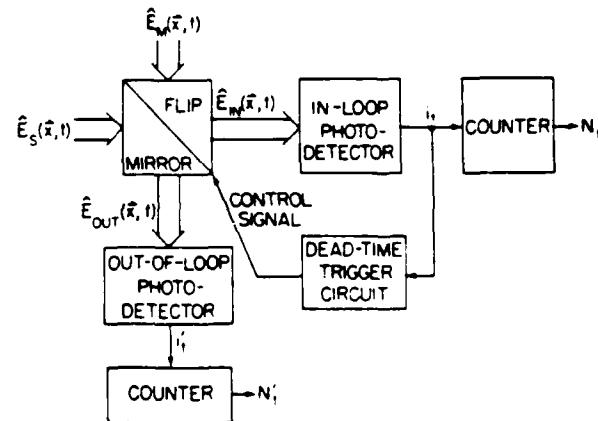
Fig. 5. Quantum-photodetection configuration that can be used to represent the DTMPP experiments carried out by Walker and Jakeman.[19]
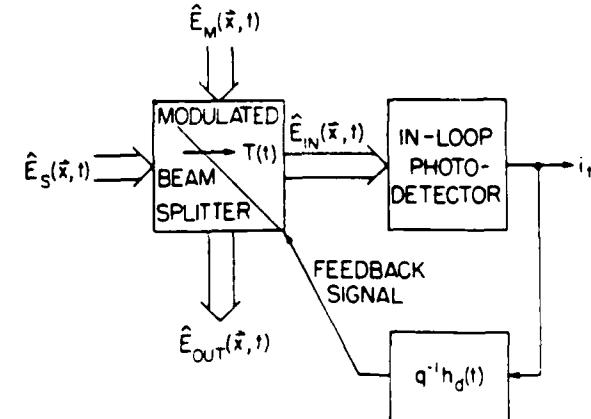

Fig. 6. Modulated beam-splitter version of the flip-mirror subsystem from Fig. 5, the beam splitter's intensity transmission obeys $T(t) = \sum_i h_d(t - t_i)$ with $h_d(t) = 1$ for $0 < t < \tau_d$.

$$\hat{E}_{OUT}(\mathbf{x}, t) = -[1 - T(t)]^{1/2}\hat{E}_S(\mathbf{x}, t) + [T(t)]^{1/2}\hat{E}_M(\mathbf{x}, t), \quad (37)$$

with beam-splitter transmission given by

$$T(t) = 1 - q^{-1} \int i_t h_d(t - \tau)d\tau \quad (38)$$

in terms of the causal linear system impulse response

$$h_d(t) = \begin{cases} 1 & \text{for } 0 < t \le \tau_d \\ 0 & \text{otherwise} \end{cases}. \quad (39)$$

Because i_t has an impulse of area q at each in-loop photo-event and because such events are impossible when the in-loop detector is illuminated by the vacuum-state field $\hat{E}_M$, it is easily verified that Eqs. (36)–(39) correctly describe the flip-mirror arrangement.

To proceed further, we replace the classical in-loop and out-of-loop photocurrents i_t and i_t' by their associated operator representations [cf. Eqs. (29), (30)]:

$$\hat{i}_t = q \int_{A_d} d\mathbf{x}\hat{E}_{IN}'^{*}(\mathbf{x}, t)\hat{E}_{IN}'(\mathbf{x}, t) \quad (40)$$

and

$$\hat{i}_t' = q \int_{A_d} d\mathbf{x}\hat{E}_{OUT}'^{*}(\mathbf{x}, t)\hat{E}_{OUT}'(\mathbf{x}, t), \quad (41)$$

where

$$\hat{E}_j'(\mathbf{x}, t) = \eta^{1/2}\hat{E}_j(\mathbf{x}, t) + (1 - \eta)^{1/2}\hat{E}_{jt}(\mathbf{x}, t) \quad (42)$$

for $j =$ IN, OUT, and $\{\hat{E}_{INt}, \hat{E}_{OUTt}\}$ are vacuum-state field operators. We next employ the high-mean-count condition to justify the following linearization[17] of Eqs. (36)–(38):

$$\langle\hat{E}_{IN}\rangle \approx \langle T\rangle^{1/2}\langle\hat{E}_S\rangle = (\langle T\rangle P/h\nu_0 A_d)^{1/2}, \quad (43)$$

$$\Delta\hat{E}_{IN}(\mathbf{x}, t) \equiv \hat{E}_{IN}(\mathbf{x}, t) - \langle\hat{E}_{IN}\rangle$$
$$\approx \langle T\rangle^{1/2}\Delta\hat{E}_S(\mathbf{x}, t) + (1 - \langle T\rangle)^{1/2}\hat{E}_M(\mathbf{x}, t)$$
$$+ [\Delta\hat{T}(t)/2\langle T\rangle^{1/2}]\langle\hat{E}_S\rangle, \quad (44)$$

$$\langle\hat{E}_{OUT}\rangle \approx -(1 - \langle T\rangle)^{1/2}\langle\hat{E}_S\rangle, \quad (45)$$

$$\Delta\hat{E}_{OUT}(\mathbf{x}, t) \approx -(1 - \langle T\rangle)^{1/2}\Delta\hat{E}_S(\mathbf{x}, t) + \langle T\rangle^{1/2}$$
$$\times \hat{E}_M(\mathbf{x}, t) + [\Delta\hat{T}(t)/2(1 - \langle T\rangle)^{1/2}]\langle\hat{E}_S\rangle, \quad (46)$$

$$\langle T\rangle \approx 1 - \int \eta A_d\langle\hat{E}_{IN}\rangle^2 h_d(t - \tau)d\tau$$
$$= 1 - \eta A_d\langle\hat{E}_{IN}\rangle^2 \tau_d, \quad (47)$$

and

$$\Delta\hat{T}(t) \approx -\int 2\eta^{1/2}\langle\hat{E}_{IN}\rangle \text{Re}\left\{\int_{A_d} d\mathbf{x}[\eta^{1/2}\Delta\hat{E}_{IN}(\mathbf{x}, \tau)\right.$$
$$\left. + (1 - \eta)^{1/2}\hat{E}_{INt}(\mathbf{x}, \tau)]\right\}h_d(t - \tau)d\tau. \quad (48)$$

The equations for the mean values are easily solved, yielding

$$\langle\hat{E}_{IN}\rangle = (P/h\nu_0 A_d)^{1/2}[1 + (\eta P\tau_d/h\nu_0)]^{1/2}, \quad (49)$$

$$\langle\hat{E}_{OUT}\rangle = -[(P/h\nu_0 A_d) - \langle\hat{E}_{IN}\rangle^2]^{1/2}, \quad (50)$$

from which it follows that

$$\langle N_t\rangle = q^{-1}\int_0^t \langle i_t\rangle d\tau \approx \eta\langle\hat{E}_{IN}\rangle^2 A_d t$$
$$\approx \lambda t/(1 + \lambda\tau_d) \quad (51)$$

and

$$\langle N_t'\rangle = q^{-1}\int_0^t \langle i_t'\rangle d\tau \approx \eta\langle\hat{E}_{OUT}\rangle^2 A_d t$$
$$\approx \lambda t - \lambda t/(1 + \lambda\tau_d), \quad (52)$$

where $\lambda = \eta P/h\nu_0$, as in Eqs. (31)–(34). Equation (51) reproduces Eq. (33), as promised. Equation (52) is, in fact, self-evident in that, by construction, $N_t + N_t'$ corresponds to performing quantum efficiency η photocounting on the coherent-state field $\hat{E}_S$. Thus $N_t + N_t'$ must be a Poisson process of rate λ. The less obvious results concern the second moments, with which we deal below.

By using the high-mean-count linearization we know that

$$\text{var}(N_t) \approx 4\eta A_d\langle\hat{E}_{IN}\rangle^2\left\langle\left[\int_0^t d\tau\Delta\hat{E}_{IN1}'(\tau)\right]^2\right\rangle, \quad (53)$$

$$\text{var}(N_t') \approx 4\eta A_d\langle\hat{E}_{OUT}\rangle^2\left\langle\left[\int_0^t d\tau\Delta\hat{E}_{OUT1}'(\tau)\right]^2\right\rangle, \quad (54)$$

and

$$\text{cov}(N_t, N_t') \approx 4\eta A_d\langle\hat{E}_{IN}\rangle\langle\hat{E}_{OUT}\rangle\left\langle\int_0^t d\tau\Delta\hat{E}_{IN1}'(\tau)\right.$$
$$\left. \times \int_0^t d\tau\Delta\hat{E}_{OUT1}'(\tau)\right\rangle, \quad (55)$$

where

$$\Delta\hat{E}_{j1}'(t) \equiv A_d^{-1/2}\int_{A_d} d\mathbf{x} \text{Re}[\hat{E}_j'(\mathbf{x}, t) - \langle\hat{E}_j'(\mathbf{x}, t)\rangle] \quad (56)$$

for $j =$ IN, OUT represents the field-quadrature fluctuations that beat with the strong mean fields to produce the photocount fluctuations. From formulas (44) and (46)–(50) we have that

$$\Delta\hat{E}_{IN1}'(t) = [\eta/(1 + \lambda\tau_d)]^{1/2}\Delta\hat{E}_{S1}(t) + [\eta\lambda\tau_d/(1 + \lambda\tau_d)]^{1/2}$$
$$\times \hat{E}_{M1}(t) + (1 - \eta)^{1/2}\hat{E}_{IN11}(t) - (\eta P/h\nu_0)$$
$$\times \int d\tau\Delta\hat{E}_{IN1}'(\tau)h_d(t - \tau) \quad (57)$$

and

$$\Delta\hat{E}_{OUT1}'(t) = -[\eta\lambda\tau_d/(1 + \lambda\tau_d)]^{1/2}\Delta\hat{E}_{S1}(t)$$
$$+ [\eta/(1 + \lambda\tau_d)]^{1/2}\hat{E}_{M1}(t) + (1 - \eta)^{1/2}$$
$$\times \hat{E}_{OUT11}(t) - [\eta P/h\nu_0(\lambda\tau_d)^{1/2}]$$
$$\times \int d\tau\Delta\hat{E}_{IN1}'(\tau)h_d(t - \tau), \quad (58)$$

where the time-dependent quadrature operators are obtained from the corresponding space–time-dependent field operators, as in Eq. (56). Equations (57) and (58) are linear-feedback forms that are easily solved by means of Fourier transformation, i.e., with

$$\Delta\tilde{E}_{IN1}'(f) \equiv \int dt \Delta\tilde{E}_{IN1}'(t)\exp(-i2\pi ft), \qquad (59)$$

etc., we can show that

$$\Delta\tilde{E}_{IN1}'(f) = \frac{[\eta/(1+\lambda\tau_d)]^{1/2}[\Delta\tilde{E}_{S1}(f)+(\lambda\tau_d)^{1/2}\tilde{E}_{M1}(f)]+(1-\eta)^{1/2}\tilde{E}_{INc1}(f)}{1+(\eta P/h\nu_0)H_d(f)} \qquad (60)$$

and

$$\Delta\tilde{E}_{OUT1}'(f) = -[\eta\lambda\tau_d/(1+\lambda\tau_d)]^{1/2}\Delta\tilde{E}_{S1}(f)$$

$$+ [\eta/(1+\lambda\tau_d)]^{1/2}\tilde{E}_{M1}(f)$$
$$+ (1-\eta)^{1/2}\tilde{E}_{OUTc1}(f)$$

$$- [\eta P/h\nu_0(\lambda\tau_d)^{1/2}]\Delta\tilde{E}_{IN1}(f)H_d(f), \qquad (61)$$

where

$$H_d(f) = \exp(-i\pi f\tau_d)[\sin(\pi f\tau_d)]/\pi f \qquad (62)$$

is the frequency response associated with the impulse response $h_d(t)$. It is now a simple matter to use the coherent-state quadrature-fluctuation statistics for $\Delta\hat{E}_{S1}$, $\hat{E}_{M1}$, $\hat{E}_{INc1}$, and $\hat{E}_{OUTc1}$ to prove that

$$\langle\Delta\hat{E}_{j1}'(t+\tau)\Delta\hat{E}_{k1}'(t)\rangle = \int df S_{E_{j1}E_{k1}}(f)\exp(i2\pi f\tau), \qquad (63)$$

for $j, k =$ IN, OUT, where the noise spectra are as follows:

$$S_{E_{IN1}E_{IN1}}(f) = 1/4|1+(\eta P/h\nu_0)H_d(f)|^2, \qquad (64)$$

$$S_{E_{OUT1}E_{OUT1}}(f) = 4^{-1} + |(\eta P/h\nu_0)H_d(f)|^2/4\lambda\tau_d|1$$
$$+ (\eta P/h\nu_0)H_d(f)|^2, \qquad (65)$$

and

$$S_{E_{IN1}E_{OUT1}}(f) = -[\eta P/h\nu_0(\lambda\tau_d)^{1/2}]H_d^*(f)S_{E_{IN1}E_{IN1}}(f). \qquad (66)$$

The behavior of the spectra [Eqs. (64) and (65)] is illustrated in Fig. 7.

To complete the second-moment analysis, we note that the high-mean-count limit requires that $t \gg \tau_d$ so that only the low-frequency behavior of the preceding noise spectra contribute to the variance and covariance in formulas (53)–(55). We then find

$$var(N_t) \approx 4[\lambda t/(1+\lambda\tau_d)]S_{E_{IN1}E_{IN1}}(0)$$

$$= \lambda t/(1+\lambda\tau_d)^3 = \langle N_t\rangle/(1+\lambda\tau_d)^2, \qquad (67)$$

$$var(N_t') \approx 4[\lambda t\lambda\tau_d/(1+\lambda\tau_d)]S_{E_{OUT1}E_{OUT1}}(0)$$

$$= [\lambda t\lambda\tau_d/(1+\lambda\tau_d)][1+\lambda\tau_d/(1+\lambda\tau_d)^2]$$

$$= \langle N_t'\rangle[1+\lambda\tau_d/(1+\lambda\tau_d)^2], \qquad (68)$$

and

$$cov(N_t, N_t') \approx 4[\lambda t(\lambda\tau_d)^{1/2}/(1+\lambda\tau_d)]S_{E_{IN1}E_{OUT1}}(0)$$

$$= \lambda t\lambda\tau_d/(1+\lambda\tau_d) \qquad (69)$$

Formula (67) reproduces the semiclassical dead-time result [Eq. (34)]; it is sub-Poisson for all τ_d values. Formulas (68) and (69) are novel. In dead-time variance, formula (68) is

the variance of the counts missed because of dead time; it is super-Poisson for all $\lambda\tau_d$ values. As a consistency check on formulas (68) and (69), we note that they imply that

$$var(N_t + N_t') = var(N_t) + var(N_t') + 2\,cov(N_t, N_t') = \lambda t. \qquad (70)$$

in agreement with the previously stated physical argument that $N_t + N_t'$ must be a Poisson process of rate λ. Moreover, even though we gave a quantum version of that argument, $N_t + N_t'$ being a Poisson process of rate λ also follows from the semiclassical shot-noise descriptions.[39]

Before proceeding to the NLFP case, some elaboration on what has just been shown is in order. Both the semiclassical and the quantum theories of photodetection predict the same mean and variance for the DTMPP in-loop count record N_t. This agreement is not coincidental. The semiclassical statistics will be the same as those of the quantum theory in any closed-loop photodetection arrangement in which breaking the feedback loop leaves the photodetector


(a)

(b)
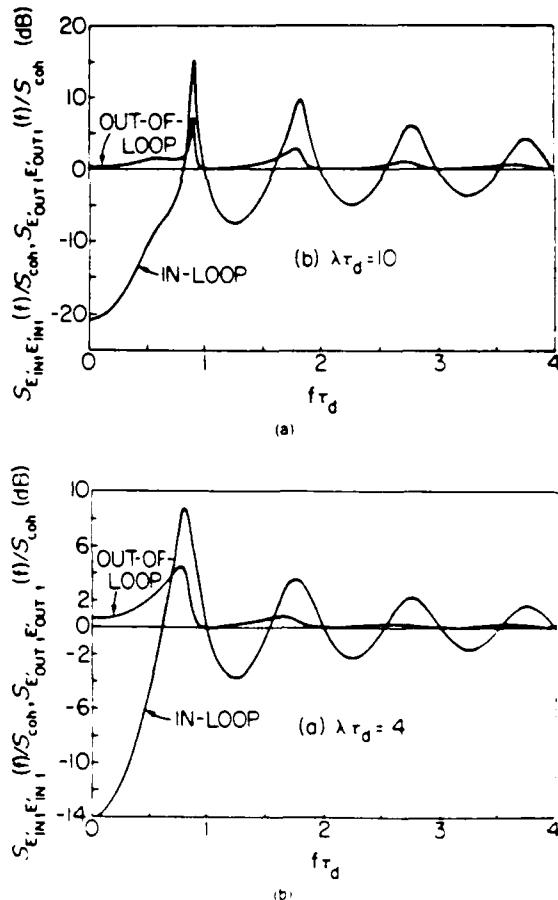
Fig. 7. Normalized first-quadrature noise spectra $S_{E_{IN1}E_{IN1}}$ and $S_{E_{OUT1}E_{OUT1}}$ (in decibels) versus normalized frequency $f\tau_d$ from the DTMPP quantum analysis. S_{coh} is the coherent-state noise spectrum. (a) $\lambda\tau_d = 4$ and (b) $\lambda\tau_d = 10$.

illuminated by a classical state [in the usual open-loop sense, cf. Eq. (23)]. In other words, despite our use of operator representations for photocurrents, the actual quantum measurement occurs at the photodetector, and the feedback path is entirely classical. Thus, if the open-loop illumination does not require the use of quantum photodetection theory, then neither does the closed-loop illumination. Conversely, if the open-loop illumination is nonclassical, then the closed-loop behavior must be analyzed quantum mechanically. For example, suppose the $\hat{E}_S$ field in the DTMPP analysis is in a broadband squeezed state with the mean field given by Eq. (35) and low-frequency first-quadrature noise squeezing[40]

$$S_{E_{S1}E_{S1}}(f) \approx \gamma/4 \qquad \text{for } |f| \leq 1/\pi\tau_d, \tag{71}$$

with $\gamma \ll 1$. Equations (67) and (68) then become

$$\text{var}(N_t) \approx \langle N_t \rangle [1 - (1 - \gamma)\eta/(1 + \lambda\tau_d)]/(1 + \lambda\tau_d)^2 \tag{72}$$

and

$$\text{var}(N_t') \approx \langle N_t' \rangle [1 + \lambda\tau_d/(1 + \lambda\tau_d)^2 - \eta\lambda\tau_d(1 - \gamma)$$
$$\times (2 + \lambda\tau_d)^2/(1 + \lambda\tau_d)^3], \tag{73}$$

respectively. For $\lambda\tau_d \gg 1$ and $\eta \approx 1$, the out-of-loop photocounts N_t' are strongly sub-Poisson, a result that cannot be obtained semiclassically as N_t' is an open-loop count record.

B. Negative-Linear-Feedback Process

Our second example of closed-loop photodetection is the NLFP whose semiclassical construct is sketched in Fig. 8. In this arrangement, the in-loop photocurrent i_t drives an optical intensity modulator through a causal linear time-invariant filter of impulse response $h_f(t)$. As a result, for nonrandom input light of constant power P_0, the optical power emerging from the intensity modulator obeys the negative-feedback law

$$P_t(\{i_\tau: \tau < t\}) = P_0 - (P_1/q)\int_{-\infty}^t i_\tau h_f(t - \tau)d\tau, \tag{74}$$

conditioned on knowledge of the in-loop event history. To ensure that Eq. (74) represents negative feedback, it is sufficient to require $P_1 \geq 0$ and $h_f(t) \geq 0$ for all t.[41] Because $P_t(\{i_\tau:\tau < t\}) \geq 0$ must prevail under all circumstances, Eq. (74) should be viewed as an approximation, even under the preceding negative-feedback conditions, whose validity requires that the feedback term on the right-hand side in Eq. (74) be smaller than the input power with overwhelming probability. It is interesting to note that the DTMPP is a NLFP with $h_f(t) = h_d(t)$ and $P_1 = P_0$. For the DTMPP no probabilistic restriction was needed to guarantee that $P_t \geq 0$, but stochastic linearization expedited the analysis. In what follows, we treat the NLFP, with an arbitrary negative feedback $h_f(t)$, using a similar stochastic linearization both to ensure $P_t \geq 0$ and to effect steady-state statistical analysis. Moreover, to make explicit the behavior of our general results, we use the single-pole filter

$$h_f(t) = \begin{cases} \exp(-t/\tau_f) & t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{75}$$

as a running example. We treat the first- and second-moment statistics of the in-loop photocurrent i_t and the out-of-

loop photocurrent i_t' using the semiclassical and quantum theories, beginning with the former.

We have from stationary-process MCR theory [see Eqs. (13) and (7)] that

$$\langle i \rangle = qw_1 = q \lim_{\Delta t \to 0} [\Delta t^{-1} \Pr(\Delta N_t = 1)] \tag{76}$$

and likewise

$$\langle i' \rangle = q \lim_{\Delta t \to 0} [\Delta t^{-1} \Pr(\Delta N_t' = 1)]. \tag{77}$$

If we employ iterated expectation on the right-hand side in Eq. (76), using the incremental SEPP description [Eqs. (3)–(5)] with $\mu_t = \eta\epsilon P_t/h\nu_0$ for ϵ the beam splitter's intensity transmission, we find that

$$\langle i \rangle = q\eta\epsilon \langle P_t \rangle /h\nu_0 = (q\eta\epsilon/h\nu_0)[P_0 - (P_1/q) \langle i \rangle H_f(0)], \tag{78}$$

where $H_f(f)$ is the frequency response associated with $h_f(t)$. Equation (78) can be solved for $\langle i \rangle$, yielding

$$\langle i \rangle = (q\eta\epsilon P_0/h\nu_0)/[1 + (\eta\epsilon P_1/h\nu_0)H_f(0)], \tag{79}$$

which reduces to

$$\langle i \rangle = (q\eta\epsilon P_0/h\nu_0)/(1 + \eta\epsilon P_1\tau_f/h\nu_0) \tag{80}$$

for the single-pole filter example. In a similar manner we can show that

$$\langle i' \rangle = q\eta(1 - \epsilon) \langle P_t \rangle /h\nu_0. \tag{81}$$

Somewhat greater effort is required to deduce the second moments, which are dealt with below.

From Eqs. (14) and (7) of the stationary-process MCR theory, it follows that the in-loop photocurrent covariance is of the form

$$\text{cov}(i_{t+\tau}, i_t) = q \langle i \rangle \delta(\tau) + \mathcal{K}_{ii}(\tau), \tag{82}$$

with a symmetric nonsingular component given by

$$\mathcal{K}_{ii}(\tau) = q^2 \lim_{\Delta t \to 0} |(\Delta t)^{-2} \Pr[\Delta N_{t+\tau}, \Delta N_t = 1]| - \langle i \rangle^2 \tag{83}$$

for $\tau \neq 0$. By iterated expectation and the incremental SEPP description, we obtain
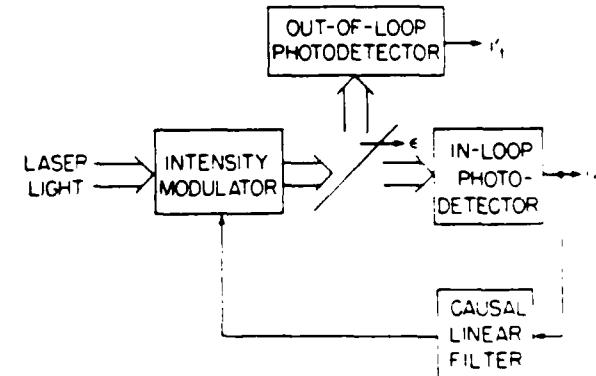
Fig. 8. Semiclassical-photodetection configuration for the NLFP experiments carried out by Machida and Yamamoto and Yamamoto.

and out-of-loop photocurrents correspond to the operator measurements, Eqs. (40) and (41), respectively, where the effective fields $\hat{E}_{IN}'$ and $\hat{E}_{OUT}'$ are obtained from Eq. (42) and the ordinary beam-splitter relations

$$\hat{E}_{IN}(\mathbf{x}, t) = \epsilon^{1/2}\hat{E}(\mathbf{x}, t) + (1 - \epsilon)^{1/2}\hat{E}_B(\mathbf{x}, t). \quad (95)$$

$$\hat{E}_{OUT}(\mathbf{x}, t) = -(1 - \epsilon)^{1/2}\hat{E}(\mathbf{x}, t) + \epsilon^{1/2}\hat{E}_B(\mathbf{x}, t). \quad (96)$$

where yet another vacuum-state field operator $\hat{E}_B$ is introduced. The introduction of the vacuum-state fields $\{\hat{E}_M, \hat{E}_B\}$ comes from the necessity of preserving open-loop commutator brackets after propagation through the beam splitters when the feedback path is absent. The vacuum-state fields $\{\hat{E}_{IN_i}, \hat{E}_{OUT_i}\}$ are required for lossy ($\eta < 1$) detectors because of the fluctuation–dissipation theorem, which, loosely stated. implies that quantum noise is injected whenever loss is encountered. again because of commutator conservation.[45] An investigation of closed-loop NLFP commutator behavior

real-quadrature fluctuations. i.e., those in phase with the mean field.

The first and second moments of i_t and i_t' are easily obtained from Eqs. (98)–(102). With the linearized formulation we know that

$$\langle i_t \rangle = qA_d(\langle \hat{E}_{IN}' \rangle)^2$$

$$= (q\eta\epsilon P_0/h\nu_0)/[1 + (\eta\epsilon P_1/h\nu_0)H_f(0)] \quad (103)$$

and

$$\langle i_t' \rangle = qA_d(\langle \hat{E}_{OUT}' \rangle)^2$$

$$= [q\eta(1 - \epsilon)P_0/h\nu_0]/[1 + (\eta\epsilon P_1/h\nu_0)H_f(0)], \quad (104)$$

in agreement with the semiclassical results [Eqs. (79) and (81)]. Moreover, by introducing Fourier transforms as in Eq. (59). we can solve the first-quadrature relations in Eqs. (100) and (101), with the following results:

$$\Delta\tilde{E}_{IN1}'(f) = \frac{[\eta\epsilon\langle T\rangle]^{1/2}\Delta\tilde{E}_{S1}(f) + [\eta\epsilon(1 - \langle T\rangle)]^{1/2}\hat{E}_{M1}(f) + [\eta(1 - \epsilon)]^{1/2}\hat{E}_{B1}(f) + (1 - \eta)^{1/2}\hat{E}_{IN_{i}1}(f)}{1 + (\eta\epsilon P_1/h\nu_0)H_i(f)} \quad (105)$$

will be of interest below, and toward that end we introduce here the auxiliary output-field relation for the modulated beam splitter:

$$\hat{E}_{AUX}(\mathbf{x}, t) = -[1 - T(t)]^{1/2}\hat{E}_S(\mathbf{x}, t) + [T(t)]^{1/2}\hat{E}_M(\mathbf{x}, t). \quad (97)$$

At this juncture. the analysis can proceed rapidly. Invoking an operator linearization of Eqs. (93)–(96). as was done for the DTMPP. and introducing time-dependent quadrature operators, in a manner similar to Eq. (56). we find that

$$\langle \hat{E}_{IN}' \rangle = (\eta\epsilon P_0/h\nu_0)A_c[1 + (\eta\epsilon P_1/h\nu_0)H_i(0)]^{1/2}, \quad (98)$$

$$\langle \hat{E}_{OUT}' \rangle = -[(1 - \epsilon)/\epsilon]^{1/2}\langle \hat{E}_{IN}' \rangle, \quad (99)$$

$$\Delta\hat{E}_{IN_1}'(t) = [\eta\epsilon\langle T\rangle]^{1/2}\Delta\hat{E}_S(t) + [\eta\epsilon(1 - \langle T\rangle)]^{1/2}$$
$$\times \hat{E}_M(t) + [\eta(1 - \epsilon)]^{1/2}\hat{E}_B(t)$$
$$+ (1 - \eta)^{1/2}\hat{E}_{IN_i}(t) - \delta_{1j}(\eta\epsilon P_1/h\nu_0)$$
$$\times \int \Delta\hat{E}_{IN_1}'(\tau)h_i(t - \tau)d\tau. \quad (100)$$

and

$$\Delta\hat{E}_{OUT}'(t) = -[\eta(1 - \epsilon)\langle T\rangle]^{1/2}\Delta\hat{E}_S(t) - [\eta(1 - \epsilon)$$
$$\times (1 - \langle T\rangle)]^{1/2}\hat{E}_M(t) + (\eta\epsilon)^{1/2}\hat{E}_B(t)$$
$$+ (1 - \eta)^{1/2}\hat{E}_{OUT_i}(t) + \delta_{1j}[\epsilon(1 - \epsilon)]^{1/2}$$
$$\times (\eta P_1/h\nu_0)\int \Delta\hat{E}_{IN_1}'(\tau)h_i(t - \tau). \quad (101)$$

In Eqs. (100) and (101)

$$\langle T\rangle = (P_1/P_0)[1 + (\eta\epsilon P_1/h\nu_0)H_i(0)] \quad (102)$$

is the average modulated beam-splitter transmission. and $j = 1, 2$ denote the real and imaginary quadratures. respectively, of $\hat{E}_{IN}'$ too. Note that the feedback affects only the

and

$$\Delta\tilde{E}_{OUT1}'(f) = -[\eta(1 - \epsilon)\langle T\rangle]^{1/2}\Delta\tilde{E}_{S1}(f)$$
$$- [\eta(1 - \epsilon)(1 - \langle T\rangle)]^{1/2}\hat{E}_{M1}(f) + (\eta\epsilon)^{1/2}$$
$$\times \tilde{E}_{B1}(f)$$
$$+ (1 - \eta)^{1/2}\tilde{E}_{OUT_i1}(f) + [\epsilon(1 - \epsilon)]^{1/2}$$
$$\times (\eta P_1/h\nu_0)\Delta\tilde{E}_{IN1}'(f)H_i(f). \quad (106)$$

From these solutions we then obtain the quadrature noise spectra [see Eq. (63)]

$$S_{E_{IN1}E_{IN1}}(f) = 1/4[1 + (\eta\epsilon P_1/h\nu_0)H_i(f)]^{-2}. \quad (107)$$

$$S_{E_{OUT1}E_{OUT1}}(f) = 4^{-1} + \epsilon(1 - \epsilon)(\eta P_1/h\nu_0)H_i(f)|^2/4[1$$
$$+ (\eta\epsilon P_1/h\nu_0)H_i(f)]^2. \quad (108)$$

and

$$S_{E_{IN1}E_{OUT1}}(f) = [\epsilon(1 - \epsilon)]^{1/2}(\eta P_1/h\nu_0) H_i^*(f)S_{E_{IN1}E_{IN1}}(f). \quad (109)$$

The quantum second-moment derivations are completed by substituting Eqs. (107)–(109) into the relations

$$S_{ii}(f) = 4q\langle i\rangle S_{E_{IN1}E_{IN1}}(f). \quad (110)$$

$$S_{i'i'}(f) = 4q\langle i'\rangle S_{E_{OUT1}E_{OUT1}}(f). \quad (111)$$

and

$$S_{ii'}(f) = -4q[\langle i\rangle\langle i'\rangle]^{1/2}S_{E_{IN1}E_{OUT1}}(f). \quad (112)$$

which follow from linearization [cf. formulas (76)]. The photocurrent spectra and cross spectrum thus obtained coincide with the semiclassical formulas. Eqs. (85)–(87).

Even though the semiclassical and quantum theories for the NLFP yield identical statistics. their physical interpretations are quite different. Consider the semiclassical

system. When the feedback path to the intensity modulator is broken. i_r and i_r' are statistically independent stationary shot-noise processes arising from the independent random atomic excitations occurring in the two detectors under steady classical illumination. Closing the loop reduces the i_r noise level at frequencies within the loop bandwidth as in any negative-feedback stabilization scheme.[46] This in-loop reduction comes about by modulating the incoming light with a filtered version of the i_r shot noise. Insofar as the out-of-loop detector is concerned, said modulation constitutes an excess noise on its illumination in the usual open-loop sense. As a result, the out-of-loop photocurrent shows super-shot-noise fluctuations within the loop bandwidth. Subtracting the normalized photocurrents precisely cancels the negative-feedback noise reduction on i_r, with the random-modulation noise increase on i_r' leading to a sum-of-shot-noise spectrum for i_-.

Quantum mechanically, the feedback loop reduces the in-loop photocurrent spectrum by squeezing, within the loop bandwidth, the quadrature fluctuations of $\hat{E}_{IN}(\mathbf{x}, t)$ that spatially match and are in phase with its mean field. Because of the π-rad phase shift that exists between the relative phases of the $\hat{E}$ and $\hat{E}_B$ contributions to $\hat{E}_{IN}$ and $\hat{E}_{OUT}$, which is a consequence of energy conservation,[47] the negative-feedback loop exacerbates the quadrature fluctuations of $\hat{E}_{OUT}(\mathbf{x}, t)$ that spatially match and are in phase with its mean field.[20,21] As a result, i_r' has a super-shot-noise spectrum within the loop bandwidth. When the photocurrents are normalized and subtracted, the noise-current operator is

$$\Delta \hat{i}_-(t) = \hat{i}_-(t) - \langle i_- \rangle = [(1 - \epsilon)/\epsilon]^{1/2} \Delta \hat{i}_r - [\epsilon/(1 - \epsilon)]^{1/2} \Delta \hat{i}_r'$$

$$= 2\{q(1 - \epsilon)\langle t \rangle/\epsilon]^{1/2} \Delta \hat{E}_{IN}'(t)$$
$$\quad + 2[q\epsilon \; (1 - \epsilon)]^{1/2} \Delta \hat{E}_{OUT}'(t)$$

$$= 2[q(\epsilon + \epsilon_0)]^{1/2}[(1 - \epsilon)^{1/2} \Delta \hat{E}_{IN}'(t)$$
$$\quad + \epsilon^{1/2} \Delta \hat{E}_{OUT}'(t)]$$

$$= 2[q(\epsilon + \epsilon_0)]^{1/2} \hat{E}_{B}(t). \tag{113}$$

from which the sum-of-shot-noises formula [Eq. (92)] follows immediately.[20,21]

C. Commutator Relations

Additional insight into closed-loop photodetection can be developed by examining the field commutators for the quantum NLFP configuration shown in Fig. 10. For notational compactness, we limit our discussion to spatially integrated time-dependent photon-unit field operators of the form

$$\hat{E}_s(t) = A_s^{-1/2} \int_{A_s} d\mathbf{x} \hat{E}_s(\mathbf{x}, t). \tag{114}$$

etc. Free fields of this type have the following commutators:

$$[\hat{E}_s(t), \hat{E}_s(t')] = 0 \tag{115}$$

and

$$[\hat{E}_s(t), \hat{E}_s^\dagger(t')] = \delta(t - t'), \tag{116}$$

which are equivalent to the quadrature operator commutators

$$[\hat{E}_s'(t), \hat{E}_s'(t')] = [\hat{E}_s''(t), \hat{E}_s''(t')] = 0 \tag{117}$$

and

$$[\hat{E}_{S1}(t), \hat{E}_{S2}(t')] = (i/2)\delta(t - t'). \tag{118}$$

When the field fluctuations are statistically stationary, the latter imply the Heisenberg uncertainty limit

$$S_{E_{S1}E_{S1}}(f)S_{E_{S2}E_{S2}}(f) \geq 1/16 \tag{119}$$

for the quadrature fluctuation spectra. Coherent-state light achieves the minimum-uncertainty product in expression (119) with equal noise strength in each quadrature; squeezed-state light achieves the minimum-uncertainty product in expression (119) with unequal noise strength in each quadrature.

Consider the uncertainty products, similar to the left-hand side member of expression (119), for the effective fields $\hat{E}_{IN}'(t)$ and $\hat{E}_{OUT}'(t)$ that drive the in-loop and out-of-loop photodetectors. We already have expressions for the first-quadrature noise spectra. The following second-quadrature noise spectra can be deduced immediately from Eqs. (100) and (101):

$$S_{E_{IN2}E_{IN2}}(f) = S_{E_{OUT2}E_{OUT2}}(f) = 1/4. \tag{120}$$

Equation (120) shows that both second-quadrature noise spectra are coherent-state results. Physically, this is so because the linearized intensity-modulation feedback does not affect the second-quadrature fluctuations as the latter are $\pi/2$ rad out of phase with the mean field. Because $S_{E_{OUT1}E_{OUT1}}(f) > 1/4$, we see that the free field $\hat{E}_{OUT}'$ obeys the usual uncertainty principle [expression (119)]. On the other hand, $S_{E_{IN1}E_{IN1}}(f) < 1/4$ prevails within the loop bandwidth, so that $\hat{E}_{IN}'$ violates the free-field uncertainty principle. If we use the frequency-domain result [Eq. (105)] and the corresponding formula for $\Delta \hat{E}_{IN2}'(f)$, we can easily show that

$$[\hat{E}_{IN}'(t), \hat{E}_{IN}'(t')] = 0 \tag{121}$$

and

$$[\hat{E}_{IN}'(t), \hat{E}_{IN}''(t')] = \int df \frac{\cos[2\pi f(t - t')]}{1 + (\eta \epsilon P_1 h r_0)H_f(f)}. \tag{122}$$

from which the Heisenberg inequality

$$S_{E_{IN1}E_{IN1}}(f)S_{E_{IN2}E_{IN2}}(f) \geq 1/16|1 + (\eta \epsilon P_1 h r_0)H_f(f)|^2 \tag{123}$$
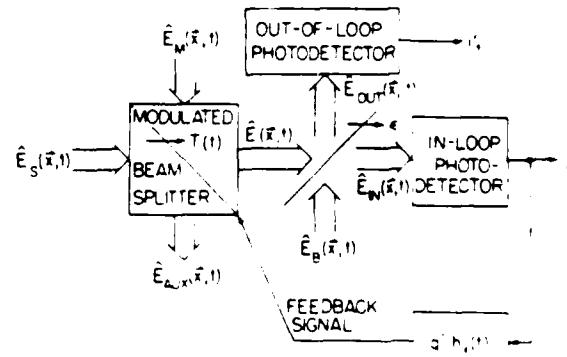
Fig. 10. Quantum photodetection configuration for the NLFP experiments performed by Machida and Yamamoto and Yamamoto.

follows readily. The in-loop effective field is operating at this uncertainty-product limit at all frequencies.

The peculiar nature of the in-loop effective-field commutator might be evidence that we have inadvertently omitted a quantum-noise contribution from our closed-loop analysis. It is germane, therefore, to examine the commutators for all the field operators involved in the arrangement shown in Fig. 10. The input operators $\{\hat{E}_S(t), \hat{E}_M(t), \hat{E}_B(t), \hat{E}_{IN_i}(t),$ $\hat{E}_{OUT_i}(t)\}$ are commuting free-field operators obeying

$$[\hat{E}_j(t), \hat{E}_{j'}(t')] = 0 \qquad (124)$$

and

$$[\hat{E}_j(t), \hat{E}_{j'}^{*}(t')] = \delta_{jj'}\delta(t - t') \qquad (125)$$

for $j, j' = S, M, B, INv, OUTv$. Whether or not the feedback loop is closed, the open-loop output operators $\{\hat{E}_{AUX}(t),$ $\hat{E}_{OUT}(t), \hat{E}_{OUT}'(t)\}$ are free fields whose commutators must also satisfy Eqs. (124) and (125). That they do so can be demonstrated by means of Eqs. (93)–(97) and the linearization procedure used to derive the quadrature-noise spectra, without recourse to additional quantum-noise sources.

The in-loop field operators $\{\hat{E}(t), \hat{E}_{IN}(t), \hat{E}_{IN}'(t)\}$ are not free fields when the loop is closed, and thus they need not have the usual commutators. We have already found the $\hat{E}_{IN}'(t)$ commutators. By similar calculations we obtain

$$[\hat{E}(t), \hat{E}(t')] = [\hat{E}_{IN}(t), \hat{E}_{IN}(t')] = 0 \qquad (126)$$

and

$$[\hat{E}(t), \hat{E}^*(t')] = [\hat{E}_{IN}(t), \hat{E}_{IN}^*(t')]$$

$$= \int df \frac{\cos[2\pi f(t - t')]}{1 + (\eta \epsilon P_1/h\nu_0)H_f(f)}. \qquad (127)$$

i.e., all three in-loop fields share the same non-free-field commutator behavior.

The special character of the in-loop commutators [Eqs. (122) and (127)] is, we believe, the hallmark of the quantum closed-loop theory. It permits the in-loop photocurrent to have a sub-shot-noise spectrum, which is the semiclassical signature of the closed loop, even though all the free-field input operators are in coherent states. Indeed, were we to have $[\hat{E}_{IN}'(t), \hat{E}_{IN}'(t')] = 0$ and $[\hat{E}_{IN}'(t), \hat{E}_{IN}'^*(t')] = \delta(t - t')$, with $\hat{E}_{IN}'(t)$ being a linear combination of only the coherent-state field operators $\{\hat{E}_S(t), \hat{E}_M(t), \hat{E}_B(t), \hat{E}_{IN_i}(t)\}$, i_t would have to be a DSPP. Its spectrum could then never reach sub-shot-noise levels.

The preceding discussion confers a special closed-loop status on the in-loop fields $\{\hat{E}(t), \hat{E}_{IN}(t), \hat{E}_{IN}'(t)\}$ in the apparatus shown in Fig. 10. This extraordinary behavior can, nevertheless, be reconciled with the obvious facts that $\hat{E}(x, t)$ represents a field in the free-space region between the modulated and ordinary beam splitters and $\hat{E}_{IN}(x, t)$ represents a field in the free-space region between the ordinary beam splitter and the in-loop photodetector. The key is optical propagation delay.

Our entire closed-loop analysis is of a lumped-element character, viz., no propagation delay whatsoever is included between the optical elements of the configuration of Fig. 10. It is well known that delay in a classical feedback loop can strongly affect its performance. Let us see what effects

optical propagation delay has on the quantum NLFP statistics and field commutators.

Suppose that there is a τ_p-sec propagation delay between the modulated beam splitter and the ordinary beam splitter and that there is no propagation delay between the ordinary beam splitter and the in-loop and out-of-loop photodetectors. In place of Eqs. (95) and (96) we then have

$$\hat{E}_{IN}(x, t) = \epsilon^{1/2}\hat{E}(x, t - \tau_p) + (1 - \epsilon)^{1/2}\hat{E}_B(x, t), \qquad (128)$$

$$\hat{E}_{OUT}(x, t) = -(1 - \epsilon)^{1/2}\hat{E}(x, t - \tau_p) + \epsilon^{1/2}\hat{E}_B(x, t). \qquad (129)$$

Mimicking the development of the no-delay case, we can show that the mean photocurrents, the photocurrent spectra, and the cross spectrum all take the forms given previously with $H_f(f)$ replaced by

$$H_f'(f) \equiv \exp(-i2\pi f\tau_p)H_f(f). \qquad (130)$$

Thus the physical discussion concluding Subsection 3.B continues to apply, insofar as moment behavior is concerned, subject to the impact of the delay factor on the right-hand side of Eq. (130) on the achievable noise squeezing. Said impact is illustrated in Fig. 11, where we have plotted $S_{ii}(f)/q(i)$ for the single-pole filter example with $\epsilon = 1/2$, $\eta P_1 \tau_f/h\nu_0 = 8$, and various τ_p/τ_f values.[48]

Now let us consider the commutator behavior when optical delay is included. Here we find that the open-loop output operators $\{\hat{E}_{AUX}(t), \hat{E}_{OUT}(t), \hat{E}_{OUT}'(t)\}$ continue to have the free-field commutators [Eqs. (124) and (125)], whereas the in-loop field operators $\{\hat{E}(t), \hat{E}_{IN}(t), \hat{E}_{IN}'(t)\}$ have the commutators given in Eqs. (121), (122), (126), and (127) with $H_f(f)$ replaced by $H_f'(f)$ from Eq. (130). If we now write

$$\int df \frac{\cos[2\pi f(t - t')]}{1 + (\eta \epsilon P_1/h\nu_0)H_f'(f)}$$

$$= \delta(t - t') - \int df \frac{(\eta \epsilon P_1/h\nu_0)H_f'(f)\cos[2\pi f(t - t')]}{1 + (\eta \epsilon P_1/h\nu_0)H_f'(f)}, \qquad (131)$$

we can use the causality of $H_f(f)$ to prove that the integral term on the right-hand side in Eq. (131) is zero for $|t - t'| < \tau_p$. Thus, for time differences smaller than the optical propagation delay, the in-loop fields have free-field commutator behavior. Physically, this means that, at any time t, the
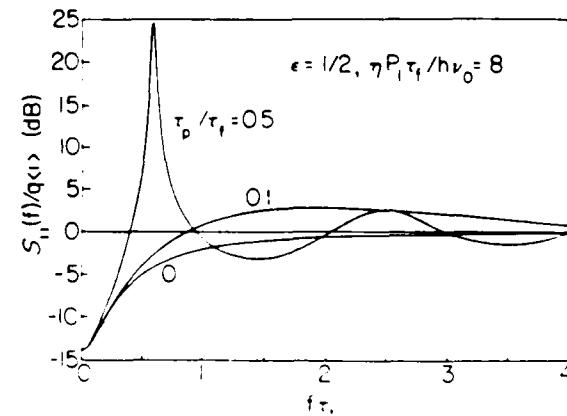
fields present in the free-space regions between the modulated beam splitter and the ordinary beam splitter and between the ordinary beam splitter and the in-loop photodetector have no special closed-loop status. It is the feedback action over time intervals in excess of the optical propagation delay that leads to the non-free-field forms in Eqs. (122) and (127). These interpretations can be tested, conceptually, by studying the transient behavior of the in-loop photocurrent statistics when the feedback path is suddenly broken. If the field that exists at any one time between the modulated beam splitter and the in-loop detector is, in essence, a classical-state free field, then breaking the feedback path at time t_0, when the loop was in its steady state, must make $\{i_j:t > t_0\}$ a DSPP. The proof is as follows.

Suppose that the closed-loop system, incorporating the τ_p-sec propagation delay, is in its statistical steady state at $t = t_0$ and that the feedback path is broken by freezing the modulated beam splitter's intensity transmission for $t \geq t_0$, i.e., by forcing

$$T(t) = \langle T \rangle - 2P_1(\eta\epsilon\langle T\rangle/Ph\nu_0)^{1/2}\int \Delta\hat{E}_{IN}'(\tau)h_j'(t_0 - \tau)d\tau$$

(132)

to prevail for $t \geq t_0$, where $h_j'(t)$ is the impulse response associated with $H_j'(f)$ and $\langle T \rangle$ is still given by Eq. (102). We then find that Eq. (100) becomes

$$\Delta\hat{E}_{IN}'(t) = [\eta\epsilon\langle T\rangle]^{1/2}\Delta\hat{E}_{S_j}(t - \tau_p) + [\eta\epsilon(1 - \langle T\rangle)]^{1/2}$$
$$\times \hat{E}_{M_j}(t - \tau_p) + [\eta(1 - \epsilon)]^{1/2}\hat{E}_{B_j}(t)$$
$$+ (1 - \eta)^{1/2}\hat{E}_{IN_0}(t) - \delta_{1j}(\eta\epsilon P_1/h\nu_0)$$
$$\times \int \Delta\hat{E}_{IN}'(\tau)h_j'(t_0 - \tau_p - \tau)d\tau$$

(133)

for $t \geq t_0, j = 1,2$. Using the remarks surrounding Eq. (131), we can show from Eq. (133) that

$$[\hat{E}_{IN}'(t), \hat{E}_{IN}'(t')] = 0$$

(134)

and

$$[\hat{E}_{IN}'(t), \hat{E}_{IN}'^\dagger(t')] = \delta(t - t')$$

(135)

hold for $t, t' \geq t_0$. Thus the effective photon-units field driving the in-loop photodetector for the $t \geq t_0$ transient regime has free-field commutator behavior. To prove $\{i_j:t > t_0\}$ is a DSPP, we need only show that $\{\hat{E}_{IN}'(t):t > t_0\}$ is in a classical state.

Using our frequency-domain solution to the steady-state feedback form [Eq. (100) with delay included], we can show that

$$\Delta\hat{E}_{IN}'(t) = \Delta\hat{E}_{EFF}'(t) - \delta_{1j}\int \Delta\hat{E}_{EFF}'(\tau)h_{EFF}(t - \tau)d\tau$$

(136)

for $j = 1, 2$ and $t \geq t_0$, where

$$\Delta\hat{E}_{EFF}'(t) = [\eta\epsilon\langle T\rangle]^{1/2}\Delta\hat{E}_{S_j}(t - \tau_p) + [\eta\epsilon(1 - \langle T\rangle)]^{1/2}$$
$$\times \hat{E}_{M_j}(t - \tau_p) + [\eta(1 - \epsilon)]^{1/2}\hat{E}_{B_j}(t)$$
$$+ (1 - \eta)^{1/2}\hat{E}_{IN_0}(t)$$

(137)

is a vacuum-state field-fluctuation operator with free-field

commutators for all times and $h_{EFF}(t)$ is the impulse response associated with the frequency response:

$$H_{EFF}(f) \equiv \frac{(\eta\epsilon P_1/h\nu_0)H_j'(f)}{1 + (\eta\epsilon P_1/h\nu_0)H_j'(f)}.$$

(138)

So, because $h_{EFF}(t)$ is zero for $t < t_p$ [cf. Eq. (131)], the integral term multiplying δ_{1j} on the right-hand side of Eq. (136) can be replaced with a classical zero-mean real-valued Gaussian random variable of variance $4^{-1}\int df/|H_{EFF}(f)|^2$. This demonstrates that $\Delta\hat{E}_{IN}'(t)$ differs from the coherent-state field operator $\Delta\hat{E}_{EFF}(t)$ on $t > t_0$ only by a c-number random variable. Hence $\{\hat{E}_{IN}(t):t > t_0\}$ is in a classical state, and $\{i_j:t > t_0\}$ is a DSPP.

4. NONCLASSICAL FIELD EXTRACTION

In this our concluding section, we address the use of nonclassical light-beam correlations to extract nonclassical open-loop fields from closed-loop photodetection systems.

A. Dead-Time-Modified Poisson Process/Photon-Twins Field Extraction

Parametric downconversion and atomic-cascade emission processes both yield the photon-twin light beams that were suggested[22,23] and recently used[25] for producing open-loop sub-Poisson light by gating procedures akin to the arrangement that we have dubbed the DTMPP.[49] A basic structure for such experiments is shown in Fig. 12. The photon-twin source produces two spatially disjoint quantum fields that illuminate identical in-loop and out-of-loop photodetectors through flip-mirror arrangements driven by the in-loop photocurrent in the manner of Fig. 5. In parametric fluorescence, the photon-units fields $\hat{E}_{REF}(\mathbf{x}, t)$ (center frequency ν_{REF}) and $\hat{E}_{SIG}(\mathbf{x}, t)$ (center frequency ν_{SIG}) are each in classical states[50]; yet, because of energy conservation at the photon-generation level, they have perfect photon-flux correlation, namely,

$$\text{var}\left\{\int_{A_j}d\mathbf{x}[\hat{E}_{SIG}^\dagger(\mathbf{x},t)\hat{E}_{SIG}(\mathbf{x},t) - \hat{E}_{REF}^\dagger(\mathbf{x},t)\hat{E}_{REF}(\mathbf{x},t)]\right\} = 0.$$

(139)
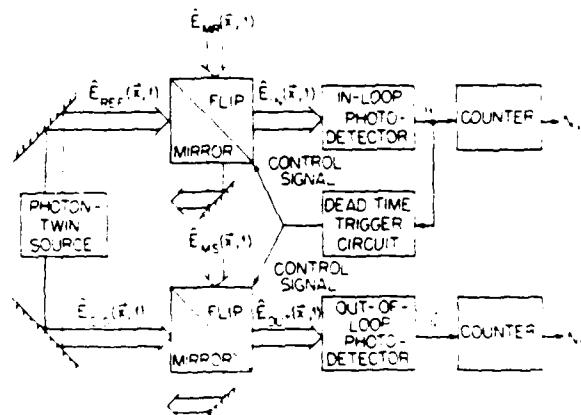
Fig. 12. Schematic for open-loop sub-Poisson beam generation by the DTMPP photon-twins experiment. The fields $\hat{E}_{SIG}$ and $\hat{E}_{REF}$ are quantum-mechanically independent vacuum-state photon-units operators at the reference and signal frequencies, respectively.

This correlation is exploited by using photodetections from the reference field to gate both the reference and the signal fields, resulting in a sub-Poisson out-of-loop photocount record, as will be shown below.

Let us assume that the detection area A_d comprises a sufficiently large number of spatial modes of th· parametric fluorescence that the low-photon coherence condition[51] applies. We can then linearize the in-loop and out-of-loop effective photon-flux operators about their mean values and show, using techniques similar to those in Subsection 3.A, that

$$\langle N_t \rangle = \langle N_t' \rangle = \lambda t/(1 + \lambda \tau_d) \tag{140}$$

and

$$\mathrm{var}(N_t) = \langle N_t \rangle/(1 + \lambda \tau_d)^2. \tag{141}$$

$$\mathrm{var}(N_t') = \langle N_t' \rangle[1 + (1 - \eta)2\lambda\tau_d(1 + \lambda\tau_d)]/(1 + \lambda\tau_d)^2. \tag{142}$$

Here, $\lambda = \eta P_{REF}/h\nu_{REF} = \eta P_{SIG}/h\nu_{SIG}$ gives the average open-loop photoemission rate in terms of the average fluorescence power P divided by the photon energy $h\nu$ at the reference and signal frequencies, and the detector quantum efficiency η was assumed to be the same at both of these frequencies. When $\lambda\tau_d \gg 1$ the out-of-loop photocount record will be sub-Poisson with a variance limited by the imperfect effective photon-flux correlation caused by the quantum-mechanically independent $\eta < 1$ quantum noise introduced at the in-loop and out-of-loop detectors, viz.,

$$\mathrm{var}(N_t') \approx \langle N_t' \rangle 2(1 - \eta). \tag{143}$$

Production of strongly sub-Poisson open-loop light by using the DTMPP photon-twins route will therefore require near-unity photodetector quantum efficiencies at both ν_{REF} and ν_{SIG}.

B. Negative-Linear-Feedback Process/Quantum Nondemolition Field Extraction

Yamamoto *et al.*[4] suggested that a Kerr-effect QND measurement[4] be used to extract a sub-shot-noise open-loop field from their feedback-modulated semiconductor-laser NLFP system. They also showed[21,44] how the back action of the QND measurement increases the second-quadrature (phase) noise on the extracted sub-shot-noise out-of-loop field, imparting to this field sufficient quantum noise to make its commutators take on the requisite free-field characteristics. These results can easily be reproduced in our NLFP construct, wherein the modulation is external to the laser, as sketched out below. The behavior of the semiconductor laser itself plays no part in our treatment.

Consider the NLFP/QND arrangement shown in Fig. 13. In this setup, two strong coherent-state fields, the center-frequency ν_s signal field $\hat{E}_s(\mathbf{x}, t)$, and the center-frequency ν_p probe field $\hat{E}_p(\mathbf{x}, t)$ are the principal inputs. The former transits a modulated beam splitter, in the manner of the ordinary (Fig. 10) NLFP experiment, before interacting with the probe field in the Kerr medium. The cross-phase-modulation interaction occurring over an Lm path in that medium engenders nonclassical coupling between the photon-flux density of the center-frequency ν_s signal field $\hat{E}_s(\mathbf{x}, t)$ and the phase of the center-frequency ν_p probe field $\hat{E}_p(\mathbf{x}, t)$ that emerge. As shown by Imoto *et al.*,[4] this quantum

correlation permits the photon-flux behavior of the signal field to be inferred from a homodyne-detection phase measurement on the probe field. This, in turn, permits the feedback loop to be closed by using the probe-field phase measurement instead of the signal-field direct-detection measurement.[21]

To make the preceding remarks explicit, let the input signal and probe fields have strong mean values

$$\langle \hat{E}_j(\mathbf{x}, t) \rangle = (P_j/h\nu_j A_d)^{1/2} \tag{144}$$

for $j = S, P$, where A_d will be regarded not only as a photodetector active area but also the cross section over which the Kerr interaction transpires. We neglect self-phase modulation, loss, and dispersion and assume that the phase shifts produced by cross-phase modulation are very small compared with a radian. In terms of the spatially integrated fields we then have

$$\hat{E}_{OUT}(t) = \{1 + i\kappa L[(P_P/h\nu_P) + 2(P_P/h\nu_P)^{1/2} \\ \times \Delta\hat{E}_{P1}(t - \tau_p)]\}\hat{E}(t - \tau_p) \tag{145}$$

and

$$\hat{E}_{IN}(t) = \{1 + i\kappa L[\langle\hat{E}\rangle^2 A_d + 2\langle\hat{E}\rangle A_d^{1/2} \\ \times \Delta\hat{E}_1(t - \tau_p)]\}\hat{E}_P(t - \tau_p), \tag{146}$$

where we have linearized about the strong mean fields. Here, κ is the cross-phase-modulation coupling constant, $\tau_p = L/c$ is the propagation delay, and subscript 1 denotes first field quadrature. The homodyne measurement on the in-loop probe field $\hat{E}_{IN}$ is arranged to sense the second field quadrature. The operator representation for the resulting homodyne photocurrent is then[5]

$$\hat{i}_t = 2q\eta(P_{LO}/h\nu_p)^{1/2}\kappa L(P_P/h\nu_p)^{1/2}[\langle\hat{E}\rangle^2 A_d \\ + 2\langle\hat{E}\rangle A_d^{1/2}\Delta\hat{E}_1(t - \tau_p)] + \hat{E}_{P2}(t - \tau_p)\} \\ + i_{vac}(t). \tag{147}$$

where P_{LO} is the local-oscillator power and i_{vac} is a zero-mean white Gaussian noise process of spectrum $q\eta(1 - \eta)P_{LO}/h\nu_p$ that represents the $\eta < 1$ quantum noise incurred in the homodyne apparatus. The feedback loop is closed by passing i_t through an amplifier of gain $K = h\nu_p/2\kappa L(P_{LO}P_P)^{1/2}$
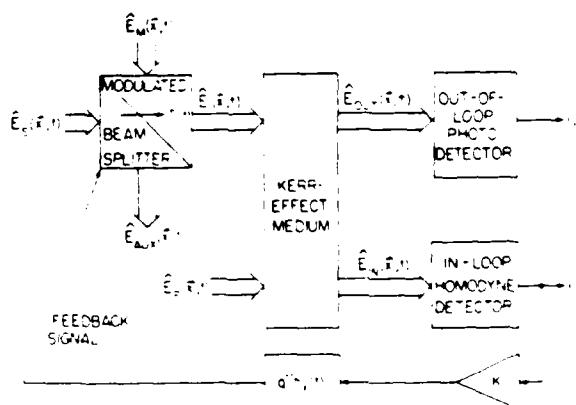

Fig. 13. Schematic for open-loop sub-shot-noise beam generation via a NLFP/QND experiment. The fields $\hat{E}_s$ and $\hat{E}_p$ are quantum-mechanically independent coherent-state photon-units operators at the signal and probe frequencies, respectively.

and the causal linear filter $q^{-1}h_f(t)$. From this point on, it is straightforward to derive the following field-quadrature and photocurrent fluctuation spectra and commutators for the out-of-loop signal field $\hat{E}_{OUT}(t)$:

$$S_{E_{OUT1}E_{OUT1}}(f)$$

$$= \frac{1 + |\eta P_i H_i'(f)/h\nu_S|^2/4(\kappa L)^2 \langle T \rangle \eta (P_P/h\nu_P)(P_S/h\nu_S)}{4|1 + (\eta P_i/h\nu_S)H_i'(f)|^2},$$

$$(148)$$

$$S_{E_{OUT2}E_{OUT2}}(f) = 4^{-1}[1 + 4(\kappa L)^2 \langle T \rangle (P_P/h\nu_P)(P_S/h\nu_S)],$$

$$(149)$$

$$S_{ii}(f) = q \langle i' \rangle [4\eta S_{E_{OUT1}E_{OUT1}}(f) + (1 - \eta)],$$

$$(150)$$

$$[\hat{E}_{OUT}(t), \hat{E}_{OUT}(t')] = 0,$$

$$(151)$$

and

$$[\hat{E}_{OUT}(t), \hat{E}_{OUT}{}^\dagger(t')] = \delta(t - t'),$$

$$(152)$$

where $\langle T \rangle = (P_s/P_S)/[1 + (\eta P_i/h\nu_S)H_i'(0)]$ is the mean beam-splitter transmission.

We see from Eqs. (148) and (150) that when the probe power is sufficiently large the free-field output $\hat{E}_{OUT}$ will be first-quadrature squeezed, within the loop bandwidth, and yield a sub-shot-noise out-of-loop photocurrent spectrum over this frequency range. We see from Eq. (149) that the second-quadrature (phase) noise on this output beam has been raised above the coherent-state level by the photon-flux (first-quadrature) probe noise introduced through the Kerr interaction [cf. Eq. (145)]. Moreover, this back action exactly reclaims the appropriate free-field commutator relations [Eqs. (151) and (152)] and therefore forces the uncertainty product obtained from Eqs. (148) and (149) to satisfy formula (119). Indeed, if we adjust the probe power P_P to minimize this uncertainty product at a selected frequency f_0, we get

$$\min_{P_i} [S_{E_{OUT1}E_{OUT1}}(f_0)S_{E_{OUT2}E_{OUT2}}(f_0)]$$

$$= \frac{(1 + \eta^2 |P_i H_i'(f_0)|h\nu_S)^2}{16|1 + \eta P_i H_i'(f_0)|h\nu_S|^2} \geq 1/16. \quad (153)$$

C. Quasi-State Synthesis

The overwhelming majority of our analysis of closed-loop photodetection has dealt with linearized feedback loops. We close by returning briefly to the general nonlinear construct. Our goal is to show how the preceding closed-loop field-extraction techniques lead to an in-principle synthesis procedure for producing an open-loop quantum light beam of arbitrary prescribed direct-detection (photocount) statistics.

Consider the open-loop photodetection of a quantized light beam using a detector of unity quantum efficiency. If the density operator (state) of this beam is totally unconstrained by previous information, then the resulting photodetection event times will comprise a similarly unconstrained SEPP. Now consider the closed-loop photodetection system in Fig. 7 where once again unity quantum efficiency is assumed. Suppose that the input light is a

highly intense beam, and the optical modulator is an infinite-bandwidth electro-optic intensity modulator driven by an arbitrary causal time-varying nonlinear system run from the photocurrent. Then this photocurrent can also be an arbitrary SEPP. In other words, using the semiclassical closed-loop theory, which is valid when the input light is in a classical state, we can synthesize any SEPP by choosing the feedback function to be such that the power falling on the photodetector satisfies[52]

$$[P_t(\{t_t, N_t\})]/h\nu_0 = \mu_t, \quad (154)$$

where μ_t is the conditional rate function for the desired SEPP. By this technique, we can use coherent-state light to create a closed-loop beam whose direct-detection statistics match those of any desired (classical or nonclassical state) open-loop beam. It should be clear from Subsections 4.A and 4.B and IV.B that quantum-beam correlations of either the photon-twins or Kerr-effect QND variety will permit extraction of said closed-loop field, leading to the aforestated synthesis procedure for open-loop beams of arbitrary prescribed direct-detection statistics. It should be noted, however, that open-loop direct-detection statistics do not uniquely determine the density operator of a quantum field,[5] so we cannot say that the procedure that we laid out can synthesize arbitrary prescribed field states. Such a synthesis procedure may perhaps be attainable from closed-loop heterodyne detection, inasmuch as it is known that open-loop heterodyne detection statistics do determine the density operator.[5]

ACKNOWLEDGMENTS

B. E. A. Saleh is also with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, Wisconsin 53706.

REFERENCES AND NOTES

1. R. J. Glauber, Phys. Rev. 130, 2529 (1963); 131, 2766 (1963).
2. P. L. Kelley and W. H. Kleiner, Phys. Rev. A 136, 316 (1964).
3. J. H. Shapiro, H. P. Yuen, and J. A. Machado Mata, IEEE Trans. Inform. Theory IT-25, 179 (1979).
4. H. P. Yuen and J. H. Shapiro, IEEE Trans. Inform. Theory IT-26, 78 (1980).
5. J. H. Shapiro, IEEE J. Quantum Electron. QE-21, 237 (1985).
6. L. Mandel, Proc. Phys. Soc. London 74, 233 (1959).
7. R. M. Gagliardi and S. Karp, *Optical Communications* (Wiley, New York, 1976).
8. B. E. A. Saleh, *Photoelectron Statistics* (Springer-Verlag, Berlin, 1978).
9. H. J. Kimble, M. Dagenais, and L. Mandel, Phys. Rev. Lett. 39, 691 (1977).
10. R. Short and L. Mandel, Phys. Rev. Lett. 51, 384 (1983).
11. M. C. Teich and B. E. A. Saleh, J. Opt. Soc. Am. B 2, 275 (1985).
12. R. E. Slusher, L. W. Hollberg, B. Yurke, J. C. Mertz, and J. F. Valley, Phys. Rev. Lett. 55, 2409 (1985).
13. R. M. Shelby, M. D. Levenson, S. H. Perlmutter, R. G. DeVoe, and D. F. Walls, Phys. Rev. Lett. 57, 691 (1986).
14. L. A. Wu, H. J. Kimble, J. L. Hall, and H. Wu, Phys. Rev. Lett. 57, 2520 (1986).

15. M. W. Maeda. P. Kumar, and J. H. Shapiro, Opt. Lett. **12**, 161 (1987).

16. J. H. Shapiro, M. C. Teich, B. E. A. Saleh, P. Kumar, and G. Saplakoglu. Phys. Rev. Lett. **56**, 1136 (1986).

17. D. L. Snyder. *Random Point Processes* (Wiley, New York. 1975).

18. D. R. Cox and V. Isham, *Point Processes* (Chapman and Hall. London, 1980).

19. J. G. Walker and E. Jakeman. Proc. Soc. Photo-Opt. Instrum. Eng. **492**, 274 (1985).

20. S. Machida and Y. Yamamoto. Opt. Commun. **57**, 290 (1986).

21. Y. Yamamoto, N. Imoto, and S. Machida. Phys. Rev. A **33**, 3243 (1986).

22. E. Jakeman and J. G. Walker. Opt. Commun. **55**, 219 (1985).

23. B. E. A. Saleh and M. C. Teich, Opt. Commun. **52**, 429 (1985).

24. N. Imoto, H. A. Haus, and Y. Yamamoto. Phys. Rev. A **32**, 2287 (1985).

25. J. G. Walker and E. Jakeman. Opt. Acta **32**, 1303 (1985); J. G. Rarity, P. R. Tapster, and E. Jakeman. Opt. Commun. **62**, 201 (1987).

26. Our photodetection model suppresses all noise sources, except those fundamental to the optical-to-electrical conversion process. We employ the photon-flux-driven description of this conversion, although our analysis is not highly sensitive to this assumption. See Ref. 5 for further elaboration on these points.

27. In effect, we are saying that the optical-to-electrical conversion process has infinite bandwidth. So long as the actual photodetector bandwidth greatly exceeds that of the loop filters employed in the analysis of Section 3, this assumption is warranted.

28. Throughout what follows we alternate between considering statistics for N_i and i_l, choosing whichever viewpoint best lends itself to the argument at hand. It follows from Eq. (1) that no loss of generality ensues from this vacillation.

29. The photon-units field operator has the commutators $[\hat{E}(\mathbf{x}, t), \hat{E}(\mathbf{x}', t')] = 0$, $[\hat{E}(\mathbf{x}, t), \hat{E}^\dagger(\mathbf{x}', t')] = \delta(\mathbf{x} - \mathbf{x}')\delta(t - t')$ in the open-loop (free-field) configuration. Closed-loop commutator behavior is treated in Section 3.

30. M. Sargent, M. O. Scully, and W. E. Lamb, *Laser Physics* (Addison-Wesley, Reading, Mass., 1974), pp. 301–303.

31. M. C. Teich and B. E. A. Saleh, Opt. Lett. **7**, 365 (1982); J. Peřina, B. E. A. Saleh, and M. C. Teich, Opt. Commun. **48**, 212 (1983).

32. M. C. Teich, B. E. A. Saleh, and J. Peřina, J. Opt. Soc. Am. B **1**, 366 (1984).

33. It follows from Theorem 1 of Ref. 4 that the statistics of the classical photocurrent i, coincide with those of the quantum measurement i_l, over any time interval. Thus, in photocurrent signal-processing calculations we may use the classical and quantum descriptors interchangeably, as demonstrated for open-loop coherent detection in Refs. 4 and 5. This same duality of viewpoint will be employed below for closed-loop configurations.

34. L. M. Ricciardi and F. Esposito. Kybernetik **3**, 148 (1966).

35. B. I. Cantor and M. C. Teich, J. Opt. Soc. Am. **65**, 786 (1975).

36. M. C. Teich and G. Vannucci, J. Opt. Soc. Am. **68**, 1338 (1978).

37. This is a stochastic linearization, i.e., the standard deviations of the filtered operator-valued fluctuations are taken to be small compared with their associated mean values. Strictly speaking, this linearization should be performed on the space-time inte-

grated-field operators used below in formulas (53)–(55). It is simpler, however, to do the linearization at the outset, as the same final results are obtained.

38. J. W. Müller. International Bureau of Weights and Measures, Pavillon de Breteuil, Sevrès, France (personal communication).

39. If we tie the two photodetectors together so that i_l drives the flip mirror but $i_l + i_{l'}$ drives the counter circuit, it is clear that $i_l + i_{l'}$ is equivalent to the photocurrent from a single photodetector under constant illumination at power P. The associated counting process for that equivalent detector in semiclassical theory is a Poisson process of rate λ.

40. The noise-reduction parameter γ cannot be made arbitrarily small at fixed mean field without violating the high-mean-count linearization assumption.

41. Negative feedback certainly prevails at all times when $P_1 h_l(t) \geq 0$ for all t, but this condition is overly restrictive. A weaker sufficient condition for the validity of the NLFP results that we develop is that the closed loop be stable, i.e., $|1 + (\eta \epsilon P_1 h\nu_0)H_l(f)| > 0$ for all f, where H_l is the frequency response associated with h_l.

42. H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York. 1968), Chap. 6.

43. That Eqs. (89)–(91) satisfy Eqs. (82)–(88) can be verified by substitution, using the causality of $H_l(f)$ and the associated feedback form $1/[1 + (\eta \epsilon P_1/h\nu_0)H_l(f)]$.

44. H. A. Haus and Y. Yamamoto. Phys. Rev. A **34**, 270 (1986).

45. W. H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, New York, 1973).

46. In particular, the i_l noise level is reduced at all frequencies for which the feedback is negative, viz., $|1 + (\eta \epsilon P_1/h\nu_0)H_l(f)| > 1$. Note that positive feedback, i.e., $|1 + (\eta \epsilon P_1/h\nu_0)H_l(f)| < 1$, can prevail at some frequencies, even though $P_1 h_l(t) \geq 0$ for all t; see Fig. 7.

47. B. L. Schumaker, Opt. Lett. **9**, 189 (1984).

48. By using $H_l'(f) = \exp[-i2\pi(\tau_p + \tau_r)]H_l(f)$ in Eq. (130) we can account for any electrical delay τ_r in the feedback path. Figure 11 can then be viewed as a function of the normalized total delay encountered when the loop is closed around the modulated beam splitter.

49. E. Jakeman and J. H. Jefferson. Opt. Acta **33**, 557 (1986).

50. The individual classical-state character of $\hat{E}_{REF}$ and $\hat{E}_{SIG}$ is easily demonstrated for a parametric downconverter in which the pump field and the nonlinear medium are treated classically, see, e.g., W. H. Louisell, *Radiation and Noise in Quantum Electronics* (McGraw-Hill, New York, 1964), pp. 274–278.

51. R. S. Kennedy, Proc. IEEE **58**, 1651 (1970).

52. Because μ, depends only on the photocount record up to time t Eq. (154) can be achieved with a causal feedback system. However, the conditional rate can diverge, see, e.g., the number-state example in Ref. 3. Thus, because our intensity modulator can act only as a variable attenuator, we must have available a very intense input beam. Furthermore, because μ, undergoes in general, abrupt transitions as each photodetection event occurs, the feedback system and the intensity modulator must possess enormous (near-infinite) bandwidths. While these consider ations may detract from the practicality of achieving Eq. (154), they do not detract from the 1:1 relationship that Eq. (154) establishes between open-loop quantum photodetection and closed-loop semiclassical photodetection.

G. Saplakoglu

G. Saplakoglu was born in Ankara. Tur-
key, on May 3, 1961. He received the
B.S. and M.S degrees in electrical engi-
neering from the Middle East Technical
University, Ankara, Turkey, in 1982 and
1984, respectively. He was awarded a
NATO science scholarship in 1984
Since September 1984 he has been work-
ing toward a Ph.D degree in electrical
engineering at the Massachusetts Insti-
tute of Technology. His research inter-
ests lie in the areas of communication
theory, optical communications, and
quantum optics.

S.-T. Ho

S.-T. Ho was born in Singapore on May
15, 1958. After Pre-University, he
served two and a half years' National
Service as an infantry officer. He re-
ceived the S.B degrees in physics and
electrical engineering in 1984, the S.M
degree in electrical engineering and com-
puter science in 1984, and the E.E. de-
gree in 1986, all from the Massachusetts
Institute of Technology. He was affili-
ated with Northrop Corporation. Nor-
wood, Massachusetts, as a student intern
from 1981 to 1983. He is now working
toward the Ph.D degree in electrical engineering at MIT. He ha-
published several papers on reducing frequency noise in semicon-
ductor lasers and on the quantum theory of atom–field interactions.
He was awarded the Newport Research Award by the Optical Soci-
ety of America in 1986. His current interest is in the generation of
the squeezed states of light. Mr. Ho is a member of Phi Beta
Kappa. Sigma Xi, Sigma Pi Sigma, and Tau Beta Pi.

Bahaa E. A. Saleh

Banaa E. A. Saleh received the B.S. de-
gree from Cairo University in 1966 and
the Ph.D. degree from Johns Hopkins
University in 1971, both in electrical en-
gineering. From 1971 to 1974 he was an
assistant professor at the University of
Santa Catarina, Brazil. Thereafter, he
joined the Max-Planck Institute. Got-
tingen, Federal Republic of Germany,
where he was involved in research in la-
ser light scattering and photon-correla-
tion spectroscopy. He is currently pro-
fessor of electrical and computer engi-
neering at the University of Wisconsin, Madison, where he has been

since 1977. He held visiting appointments at the University of
Kuwait in 1976, the School of Optometry of the University of Cali-
fornia at Berkeley in 1977, and the Columbia Radiation Laboratory
of Columbia University, New York, in 1983. He is currently in-
volved in research in image processing, optical information process-
ing, statistical optics, quantum optics, optical communication, and
vision. He is the author of *Photoelectron Statistics* (Springer, New
York, 1978) and a coeditor of *Transformations in Optical Signal
Processing* (Society of Photo-Optical, 1981). From 1980 to 1983 he
was an associate editor of the *Journal of the Optical Society of
America*, and since 1983 he has been a topical editor of the *Journal
of the Optical Society of America A*. Dr. Saleh is a Fellow of the
Optical Society of America, a senior member of the Institute of
Electrical and Electronics Engineers, and a member of Phi Beta
Kappa and Sigma Xi. In 1984 he was appointed a Fellow of the
John Simon Guggenheim Memorial Foundation.

M. C. Teich

M. C. Teich was born in New York City.
He received the S.B degree in physics
from the Massachusetts Institute of
Technology in 1961, the M.S degree in
electrical engineering from Stanford
University in 1962, and the Ph.D degree
in quantum electronics from Cornell
University in 1966. On graduating from
Cornell, he joined the MIT Lincoln Lab
oratory, where he was engaged in work
on coherent infrared detection. In 1967
he became a member of the faculty in the
Department of Electrical Engineering at
Columbia University, where he is now teaching and pursuing his
research interests in the areas of photonics, quantum optics, light-
wave communications, and sensory perception. He served as chair-
man of the department from 1978 to 1980. He is also a member of
the faculty in the Department of Applied Physics and Nuclear
Engineering and a member of the Columbia Radiation Laboratory,
the Center for Telecommunications Research, and the Columbia
Bioengineering Institute. Dr. Teich is a member of Sigma Xi, the
American Physical Society, the Institute of Electrical and Electron-
ics Engineers, the Acoustical Society of America, the Society for
Neuroscience, the American Association for the Advancement of
Science, and the New York Academy of Sciences. He served as a
member of the Editorial Advisory Panel for *Optics Letters* from
1977 to 1979. In 1969 he was the recipient of the IEEE Browder J.
Thompson Memorial Prize for his paper "Infrared Heterodyne De-
tection" and in 1981 he received the Citation Classic Award of
Current Contents for this work. He was appointed a Fellow of the
John Simon Guggenheim Memorial Foundation in 1973 and was
elected a Fellow of the Optical Society of America in 1983.

Approproximate photocounting statistics of shot-noise light with arbitrary spectrum

M. C. TEICH

Columbia Radiation Laboratory and Center for Telecommunications Research, Department of Electrical Engineering, Columbia University, New York, New York 10027, U.S.A.

and B. E. A. SALEH

Department of Electrical and Computer Engineering, University of Wisconsin, Madison, Wisconsin 53706, U.S.A.

Abstract. The photocounting statistics of shot-noise light have been studied extensively. Light exhibiting such statistical properties arises in photon-generation processes that involve a two-stage cascade of Poisson-based events, such as cathodoluminescence. Expressions for the photocounting distributions are complex because they depend on the overall mean of the distribution, the photodetector counting time, and the spectrum of the light. A simple two-parameter distribution, the Neyman Type-A (NTA), is shown to provide an excellent approximation to the photocounting statistics of shot-noise light with arbitrary spectral properties. The NTA distribution therefore plays the same role for shot-noise light that the negative-binomial distribution plays for chaotic light.

1. Introduction

The shot-noise-driven doubly stochastic Poisson point process (SNDP) [1, 2] and the Thomas point process (TPP) [3] are useful for describing sequences of events that arise from a cascade of two Poisson-based point processes, when the temporal effects of the cascading process are taken into account. These point processes have found use in a variety of fields, including optics [1–8], vision [4, 6, 9–11], and other disciplines ranging from biological information transmission to astrophysics [1, 3, 4, 12–14].

The SNDP is formed from a two-stage Poisson cascade, in which each event of a primary Poisson point process produces a virtual inhomogeneous rate function, which in turn is the source of a secondary non-stationary Poisson point process. The secondary processes are superimposed to form the overall (stationary) point process. The SNDP is a doubly stochastic Poisson point process (DSPP); it is also a special case of the Neyman Scott cluster process [1, 12]. The primary events are excluded from the overall point process in the SNDP; the TPP is distinguished only by the inclusion of the primary events. It has been shown that the SNDP counting distribution reduces to the simple Neyman Type-A (NTA) [1, 4, 15] when cascading occurs instantaneously; the TPP counting distribution similarly reduces to the Thomas (which is similar to the NTA) in the absence of time dynamics [3, 16].

In the context of optics, the theory has been used to characterize the photocounting statistics of shot-noise light. Light exhibiting such statistical

characteristics arises when the photon-generation process involves a two-stage cascade of Poisson-based events. Examples include cathodoluminescence, optical fluorescence initiated by ionizing radiation, and X-ray radiography. In processes such as these, both the impinging excitations (primary events) and the photon emissions (secondary events) are often endowed with Poisson statistics. The presence of Cerenkov photons can be accommodated by including the primary events [3].

Unfortunately, expressions for the photocounting distributions associated with shot-noise light are complex because they depend on the overall mean of the distribution $\langle n \rangle$, the counting time T and a function $h(t)$. This function is an inhomogeneous rate that describes the time course of secondary-event production upon the arrival of a primary event, and it determines the spectrum of the light. The principal parameters that characterize $h(t)$ are the average number of secondary events per primary event α (the area of $h(t)$) and the characteristic secondary-event cluster time τ_p (the width of $h(t)$).

The purpose of this paper is to investigate the validity of the Neyman Type-A distribution as an approximation for the photocounting statistics of shot-noise light with arbitrary spectrum. It is a simple distribution characterized by two parameters: its mean $\langle n \rangle$ and a multiplication parameter a. The choice of this particular distribution follows from theoretical studies demonstrating that the SNDP and TPP counting distributions approach the NTA in various limits [1, 3]. Our principal interest is in the region $\beta = T/\tau_p \approx 1$, where the specific form of $h(t)$ is likely to be important. The counting time T is then of the order of the characteristic splay-out time of the secondary pulses (τ_p) so that the validity of the approximation must be determined numerically. The NTA approximation turns out to be excellent when $h(t)$ assumes two common forms (rectangular and exponential), for all regions of the parameters α and β. Our results suggest that the approximation is probably quite suitable for arbitrary $h(t)$, α and β.

The approach we adopt is similar to that used by Bédard, Chang and Mandel [17] and by Peřina and his coworkers [18 21]. The former study dealt with the use of the negative-binomial distribution as an approximation for the photocounting statistics of chaotic light with arbitrary spectrum. The latter studies were concerned with the use of the non-central negative binomial (Laguerre) distribution as an approximation for the photocounting statistics of mixed coherent and chaotic light [22].

2. The NTA counting distribution

The NTA distributed random variable n, of mean $\langle n \rangle$ and multiplication parameter a (average number of secondary events per primary event), has a moment-generating function given by [1, 15]

$$Q_n(s) \equiv \langle \exp(-sn) \rangle = \exp \left\{ \frac{\langle n \rangle}{a} [\exp[a(\exp(-s) - 1)] - 1] \right\}. \tag{1}$$

Its variance is

$$\text{Var}(n) = (1 + a)\langle n \rangle. \tag{2}$$

The usual expression for the NTA counting distribution $p(n)$ takes the form of an infinite series [4], with a recurrence relation of the form [1]

$$(n+1)p(n+1) = \langle n \rangle \sum_{j=0}^{n} \exp(-a) \frac{a^j}{j!} p(n-j), \quad \Bigg\}$$

$$p(0) = \exp\left[\frac{\langle n \rangle}{a}(\exp(-a)-1)\right]. \quad \Bigg\} \tag{3}$$

When $a \to 0$, (1) (3) reduce to expressions for the Poisson distribution.

3. Comparison of the NTA and SNDP counting distributions

The NTA is considered a candidate for approximating the SNDP and TPP counting distributions because it is known from theoretical arguments that it approaches these complex counting distributions in important limits. For vanishingly small α, or for vanishingly small β and $\beta\alpha$, it has been shown that the SNDP counting distribution approaches the Poisson distribution for arbitrary $h(t)$ [1, 2]. This is because the secondary pulse clusters are then either very sparse ($\alpha \ll 1$) or are cut apart by the small value of T ($\beta \ll 1$). Since the NTA approaches the Poisson distribution for vanishingly small a, it certainly is a good approximation to the SNDP in this region. At the opposite extreme, when $\beta \gg 1$, the SNDP counting distribution also approaches the NTA for arbitrary $h(t)$ [1, 2]. This is because all of the secondary pulses are then captured in the long counting time T since $T \gg \tau_p$.

The comparison is effected by matching the NTA mean and variance (2) to the mean and variance of the SNDP counting distribution, which is given by

$$\mathrm{Var}(n) = \left(1 + \frac{\alpha}{\mathcal{M}}\right)\langle n \rangle. \tag{4}$$

The quantity $\mathcal{M}$ is a degrees-of-freedom parameter that is determined by $h(t)$ and T. When $h(t)$ has the rectangular form

$$h(t) = \frac{\alpha}{\tau_p} \quad (0 \leqslant t \leqslant \tau_p), \tag{5}$$

the degrees-of-freedom parameter $\mathcal{M}(\mathrm{SNDP}, \mathrm{rect})$ becomes [1]

$$\mathcal{M}(\mathrm{SNDP}, \mathrm{rect}) = \begin{cases} (\beta - \frac{1}{3}\beta^2)^{-1}, & (\beta \leqslant 1), \\ \beta(\beta - \frac{1}{3})^{-1}, & (\beta \geqslant 1). \end{cases} \tag{6}$$

The usefulness of the NTA as an approximation to the SNDP counting distribution with spectral properties determined by the rectangular form of $h(t)$ is illustrated graphically in figure 1. Six representative SNDP counting distributions $p(n)$ are shown versus count number n for various values of the SNDP multiplication parameter α and the capture ratio $\beta = T/\tau_p$. These distributions were calculated from the recursion relations derived earlier [1]. The mean count $\langle n \rangle$ is 10 in all cases. Also shown in figure 1 are six NTA distributions whose mean and variance have been matched to those of the SNDP counting distributions by equating (2) and (4), and by using (6) to solve for a in terms of α and β. In several panels of figure 1, the two curves are so nearly identical that they look like a single curve.

In figures 1 (a)–(c), α is chosen to be unity, whereas in figures 1 (d)–(f), $\alpha = 10$. $\beta = 0.05$ at the top panels, increasing to 5.0 at the bottom panels. It is apparent from
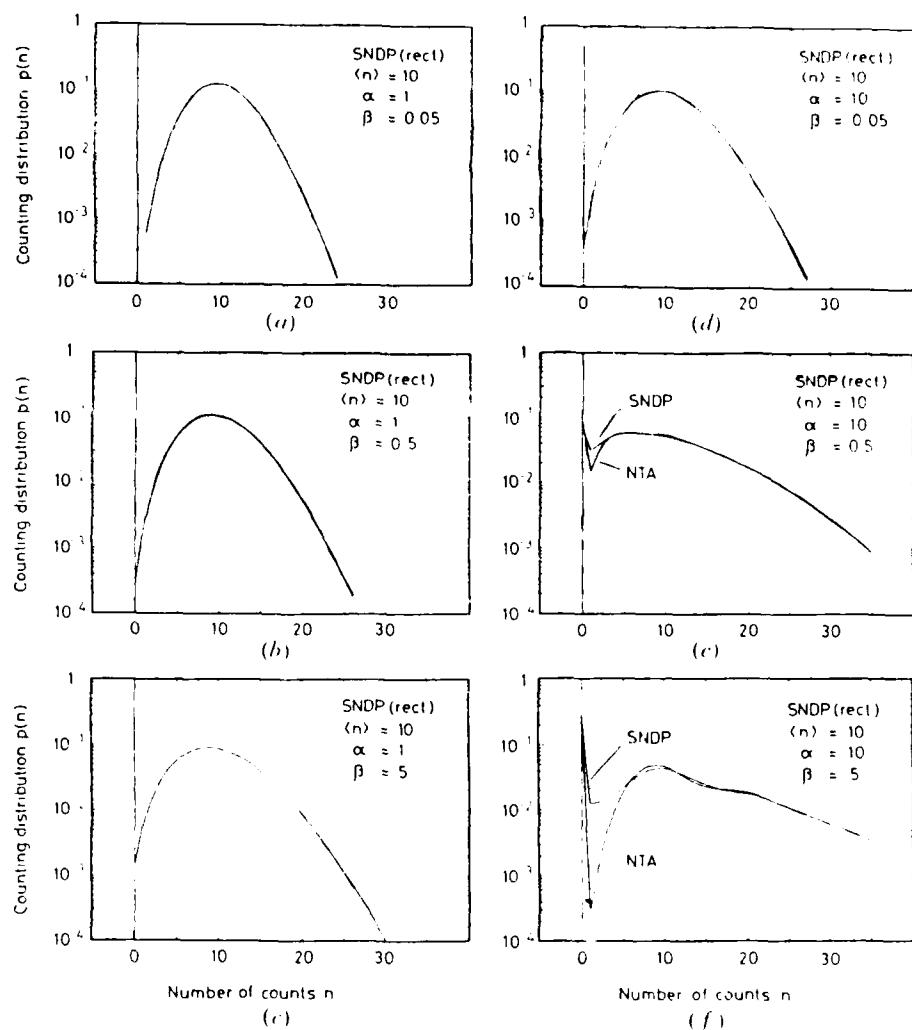
Figure 1. Six representative rectangular SNDP counting distributions compared with NTA distibutions of the same mean ($\langle n \rangle = 10$) and variance. The multiplication parameter α and the time parameter β are indicated in each panel.

figure 1 that NTA and rectangular SNDP distributions of the same mean and variance are indistinguishable for $\alpha = 1$ regardless of the value of β. The same is true for $\alpha = 10$ and $\beta = 0.05$, as in figure 1 (d); this is expected because the SNDP distribution reduces to the NTA when $h(t)$ is rectangular in shape and the counting time is very short ($\beta \ll 1$) [1]. However, the NTA provides a less-than-perfect approximation for certain ranges of the count number n, when $\alpha = 10$ and $\beta = 0.5$ and 5.

Similar results are shown in figure 2 when $h(t)$ assumes the exponential form

$$h(t) = \frac{2\alpha}{\tau_p} \exp\left(\frac{-2t}{\tau_p}\right) \quad (t \geqslant 0), \tag{7}$$
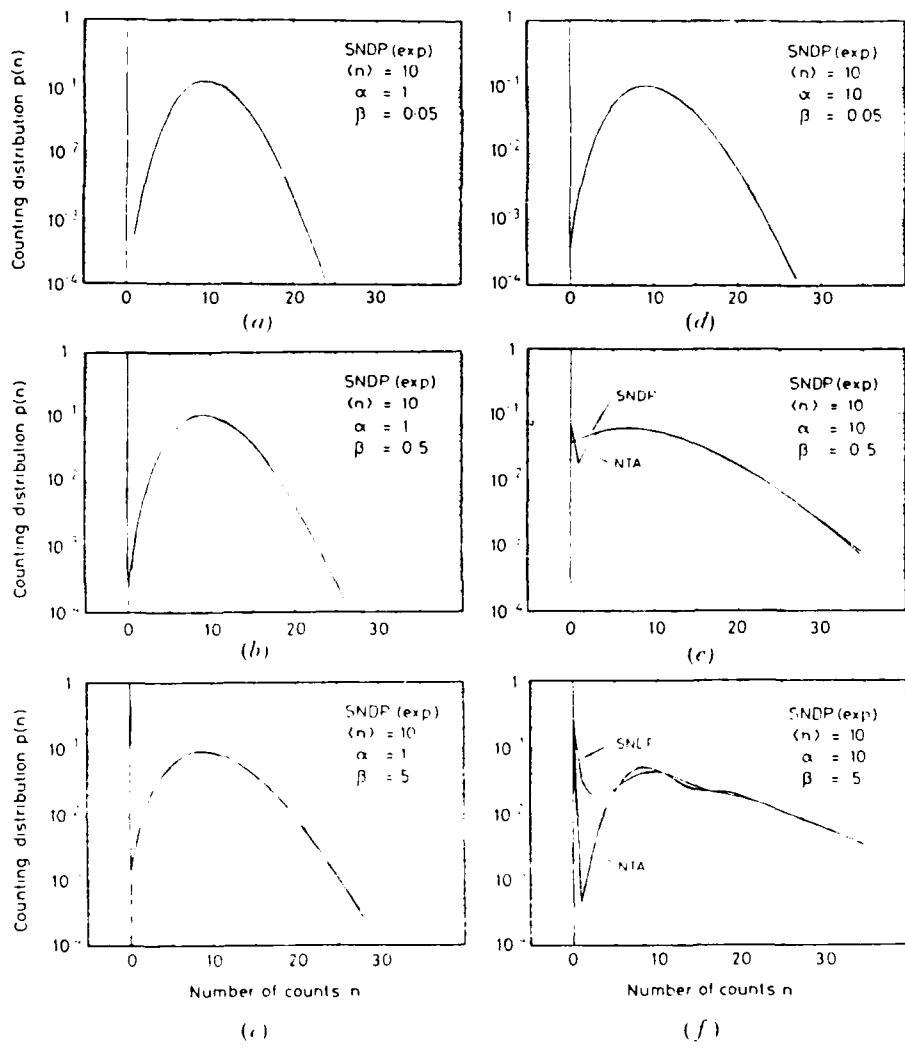
Figure 2. Six representative exponential SNDP counting distributions compared with NTA distributions of the same mean ($\langle n \rangle = 10$) and variance. The parameters α and β are indicated.

in which case the exponential degrees-of-freedom parameter $\mathcal{M}(\text{SNDP, exp})$ is given by [1]

$$\mathcal{M}(\text{SNDP}, \exp) = \frac{2\beta}{\exp(-2\beta) + 2\beta - 1}.$$ (8)

Again, various values of α and β are represented in the individual panels of figure 2, and $\langle n \rangle = 10$. The matching of the variance is again accomplished by equating (2) and (4), but now (8) is used to give a in terms of α and β.

In figures 2(a)–(c), $\alpha = 1$, whereas in figures 2(d)–(f), $\alpha = 10$. β increases from 0·05 at the top panels to 5·0 at the bottom panels. Apparently the NTA and exponential SNDP distributions are also indistinguishable (within the graphical accuracy of the figures) when $\alpha = 1$, independent of the value of β. In this case, however, there is a bit more disparity between the NTA and the exponential SNDP

distribution when $\alpha = 10$ and $\beta = 0.05$ (figure 2 (d)); this is expected [1]. In fact, the NTA misses the mark by a bit for all cases in which $\alpha = 10$. Nevertheless, a comparison of figures 1 and 2 demonstrates that the NTA fares just about as well in approximating rectangular and exponential SNDP counting distributions. The SNDP distributions appear to be smoothed versions of the NTA, in which the sharp dips of $p(n)$, at low values of n, are washed out. The tails of the distributions are approximately the same.

The goodness of the approximation can be quantitatively estimated. To accomplish this, we calculate the relative error $\delta(n)$ as a function of the count number n, where $\delta(n)$ (in %) is defined as

$$\delta(n) = 100 \times \frac{|p(n, \text{SNDP}) - p(n, \text{NTA})|}{p(n, \text{SNDP})} . \tag{9}$$

$\delta(n)$ is displayed in figures $3(a, b)$ for rectangular SNDP, and in $3(c, d)$ for exponential SNDP, counting distributions. The values of α and β are indicated in the individual panels of the figure, and again $\langle n \rangle = 10$. The error is oscillatory with the count number, indicating that the approximation is better for certain values of n. Overall, the relative error is often quite small, but it can become large for particular values of n. Expectations based on theoretical arguments are borne out. The NTA
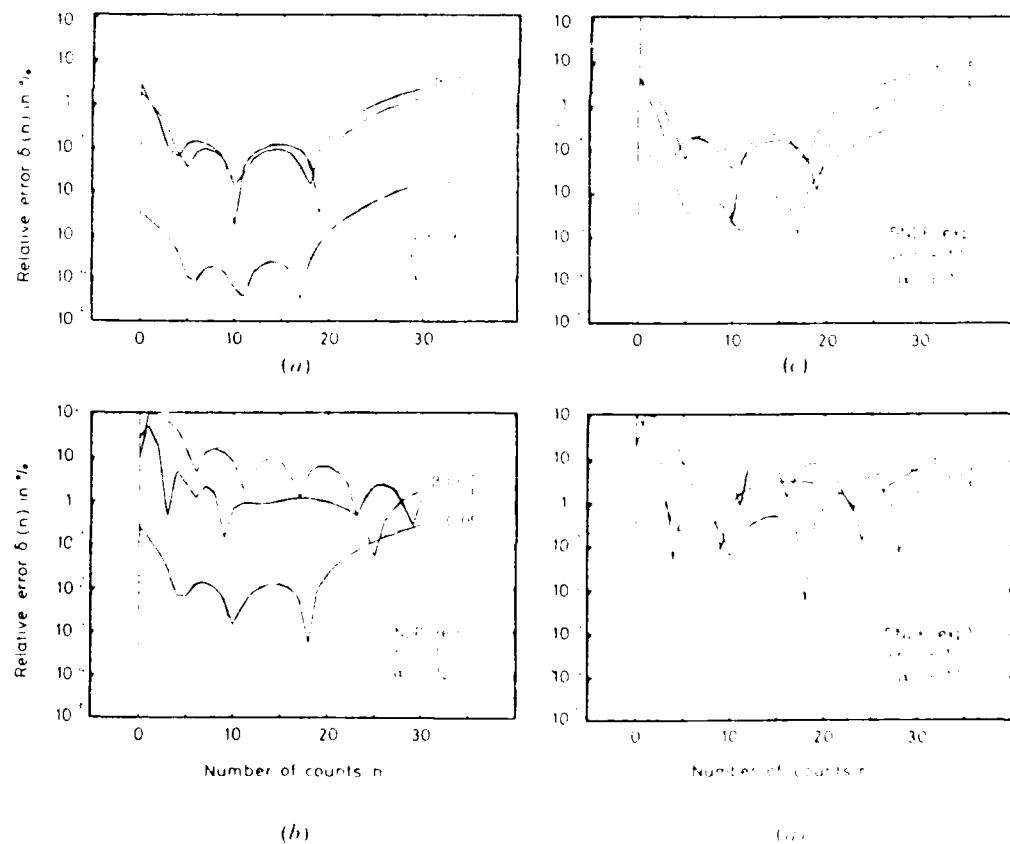

Figure 3. Relative error $\delta(n)$ (per cent) versus number of counts n for rectangular and exponential SNDP counting distributions. These curves are obtained from the distributions in figures 1 and 2. $\langle n \rangle = 10$, α and β are indicated.

provides a slightly better approximation for the rectangular SNDP distribution (i.e. the curves lie a bit lower) than for the exponential SNDP distribution. The approximation is always somewhat better for $\beta = 0.05$ than for higher values of β.

4. Comparison of the NTA and Thomas-process counting distributions

The NTA also approaches the TPP counting distribution in certain limits. The TPP counting distribution with arbitrary $h(t)$ becomes the Poisson distribution when $\alpha = 0$, or when β and $\beta\alpha$ become vanishingly small [3]. The first conditions represents a degenerate case in which there are no secondary events but Poisson primary events are carried forward, whereas the second condition represents a situation in which the secondary-pulse clusters are cut apart by the small value of T ($\beta \ll 1$). Since the NTA also approaches the Poisson when $a \to 0$, it clearly provides a good approximation to the TPP counting distribution in this limit. In the opposite case, when $\beta \gg 1$, all of the secondary events are captured in the long counting time T, along with the primary events, and the TPP counting distribution with arbitrary $h(t)$ approaches the Thomas distribution [3, 4, 16], which in turn is similar to the NTA. Finally, in the limit $\beta \gg 1$, $\alpha \gg 1$, the TPP, SNDP and NTA counting distributions all approach the fixed multiplicative Poisson distribution [4].

The variance of the TPP counting distribution is [3]

$$\mathrm{Var}\,(n) = \frac{1 + 3\alpha \mathcal{N}' + \alpha^2 \mathcal{N}}{1 + \alpha} \langle n \rangle, \tag{10}$$

where $\mathcal{N}'$ and $\mathcal{N}$ are Thomas degrees-of-freedom parameters that are determined by $h(t)$ and T. When $h(t)$ assumes the rectangular form represented in (5),

$$\mathcal{N}(\mathrm{TPP}, \mathrm{rect}) = \mathcal{N}(\mathrm{SNDP}, \mathrm{rect}), \tag{11}$$

as given by (6) and [3]

$$\mathcal{N}'(\mathrm{TPP}, \mathrm{rect}) = \begin{cases} 3(1 + \beta)^{-1} & (\beta \leqslant 1), \\ \beta(\beta - \tfrac{1}{3})^{-1} & (\beta \geqslant 1). \end{cases} \tag{12}$$

The NTA and TPP counting distributions were compared in the same way as reported earlier. The means and variances of the NTA and the TPP distributions were matched by equating (2) and (10), and by using (6), (11) and (12) to provide an expression for a in terms of α and β.

Six representative rectangular TPP counting distributions, along with six NTA distributions of the same mean and variance, are displayed in figure 4. The TPP distributions were computed from the recursion relations obtained previously [3]. The mean $\langle n \rangle$ was again chosen to be 10, and the same values of α and β were used as in figure 1 for the SNDP. In four panels of figure 4 the NTA and TPP curves are so nearly identical that they look like a single curve. Comparison of figures 1 and 4 reveals a great deal of similarity.

Again, the goodness of the approximation may be quantitatively established by use of the relative error $\delta(n)$ expressed in (9), substituting $p(n, \mathrm{TPP})$ for $p(n, \mathrm{SNDP})$. The result is displayed in figure 5. The values of $\langle n \rangle$, α and β are the same as those used for the SNDP in figures 3 (a, b). Again, the NTA approximation is somewhat better for $\beta = 0.05$ than for higher values of β. A general comparison of figures 5 and 3 quickly confirms that the NTA provides as suitable an approximation for the rectangular TPP distribution as it does for the rectangular SNDP distribution.
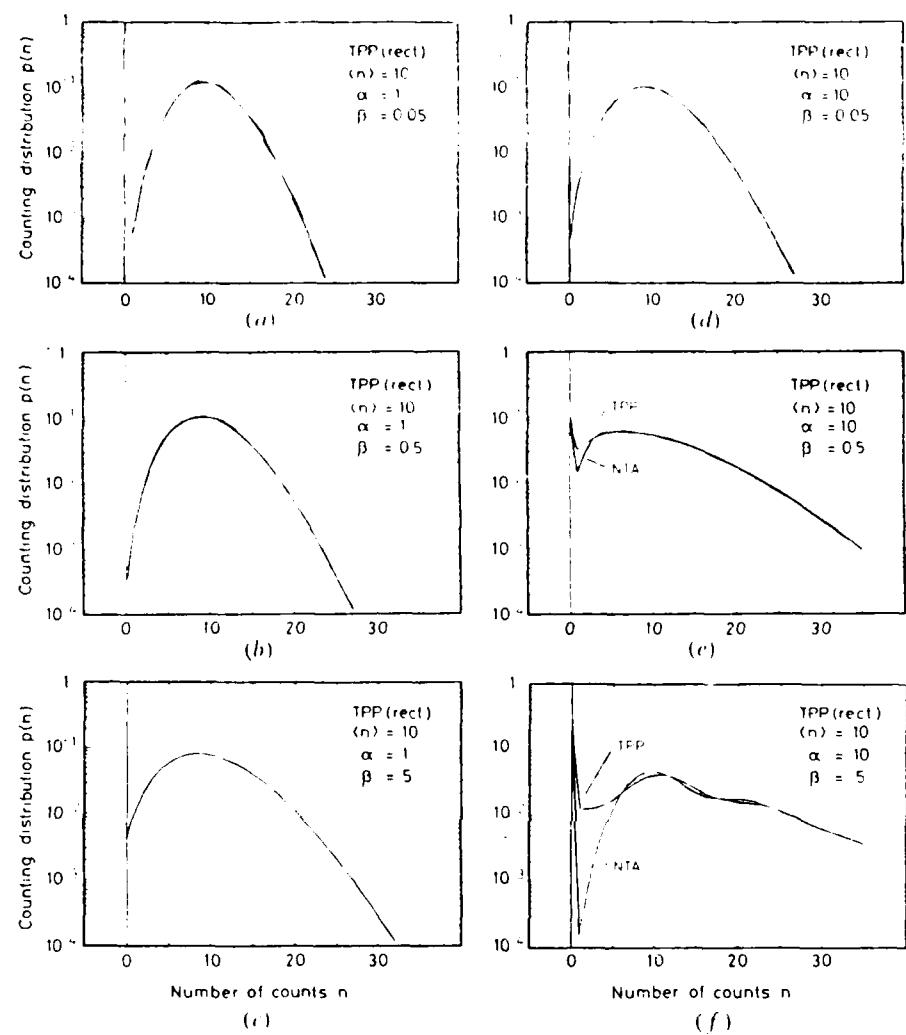
Figure 4. Six representative rectangular TPP counting distributions compared with NTA distributions of the same mean ($\langle n \rangle = 10$) and variance. The parameters α and β are indicated.

We conclude that the NTA provides an excellent approximation for the SNDP and TPP counting distributions, with rectangular and exponential impulse response functions, when the mean and variance of the distributions are matched. The NTA distribution therefore plays the same role for shot-noise light that the negative-binomial distribution plays for chaotic light.

Finally, we note that counting distributions for cascaded Poisson processes with an arbitrary number of stages have recently been calculated [23, 24]. As an extension of the work reported here, it would be of interest to examine the performance of the NTA distribution as an approximation to the cascaded-Poisson family of counting distributions. A similar approach may also be useful for approximating non-Poisson cascaded processes, such as those used for describing the generation of antibunched and sub-Poisson light [25]. Aside from the NTA, many two-stage simple time-independent counting distributions have been developed over the years.
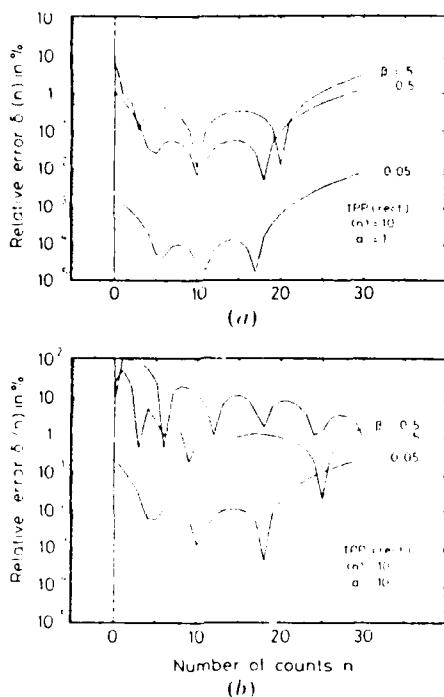
Figure 5. Relative error $\delta(n)$ (per cent) versus number of counts n for rectangular TPP counting distributions. These curves are obtained from the distributions in figure 4: $\langle n \rangle = 10$, γ and β are indicated.

Acknowledgment

References

[1] SALEH, B. E. A., and TEICH, M. C., 1982, *Proc. Inst. elect. electron. Engrs*, **70**, 229–245.
[2] TEICH, M. C., and SALEH, B. E. A., 1981, *Phys. Rev. A*. **24**, 1651 1654.
[3] MATSUO, K., TEICH, M. C., and SALEH, B. E. A., 1983, *Appl. Opt.*, **22**, 1898 1909.
[4] TEICH, M. C., 1981, *Appl. Opt.*, **20**, 2457 2467.
[5] SALEH, B. E. A., TAVOLACCI, J. T., and TEICH, M. C., 1981, *IEEE Jl quant. Electron.*, **17**, 2341 2350.
[6] SALEH, B. E. A., and TEICH, M. C., 1983, *IEEE Trans. Inf. Theory*, **29**, 939–941.
[7] SALEH, B. E. A., STOLER, D., and TEICH, M. C., 1983, *Phys. Rev. A*, **27**, 360 374.
[8] HELSTROM, C. W., and RICE, S. O., 1984, *J. comput. Phys.*, **54**, 289 324.
[9] TEICH, M. C., and SALEH, B. E. A., 1981, *J. opt. Soc. Am.*, **71**, 771 776.
[10] TEICH, M. C., PRUCNAL, P. R., VANNUCCI, G., BRETON, M. E., and MCGILL, W. J., 1982, *J. opt. Soc. Am.*, **72**, 419 431.
[11] SALEH, B. E. A., and TEICH, M. C., 1985, *Biol. Cybernet.*, **52**, 101 107.
[12] NEYMAN, J., and SCOTT, E. L., 1958, *Jl R. statist. Soc.* B, **20**, 1 43
[13] NEYMAN, J., and PURI, P. S., 1976, *Proc. natl Acad. Sci. U S A*, **73**, 3360 3363
[14] LAWRENCE, A. J., 1972, *Stochastic Point Processes Statistical Analysis, Theory, and Applications*, edited by P. A. W. Lewis (New York: Wiley-Interscience), pp. 199 256
[15] NEYMAN, J., 1939, *Ann. math. statist.*, **10**, 35 57.
[16] THOMAS, M., 1949, *Biometrika*, **36**, 18 25.

[17] Bédard, G., Chang, J. C., and Mandel, L., 1967, *Phys. Rev.*, **160**, 1496-1500.
[18] Peřina, J., 1967, *Phys. Lett.*, **24A**, 333-334.
[19] Peřina, J., Peřinová, V., Lachs, G., and Braunerová, Z., 1973, *Czech. J. Phys.* B, **23**, 1008-1013.
[20] Peřina, J., 1984, *Quantum Statistics of Linear and Nonlinear Optical Phenomena* (Dordrecht: Reidel), Chap. 5, p. 148.
[21] Hradil, Z., and Peřina, J., 1986, *J. opt. Soc. Am.* A, **3**, 1328-1333.
[22] Saleh, B. E. A., 1978, *Photoelectron Statistics* (Berlin: Springer).
[23] Matsuo, K., Saleh, B. E. A., and Teich, M. C., 1982, *J. math. Phys.*, **23**, 2353-2364.
[24] Matsuo, K., Teich, M. C., and Saleh, B. E. A., 1984, *J. math. Phys.*, **25**, 2174-2185.
[25] Teich, M. C., Saleh, B. E. A., and Peřina, J., 1984, *J. opt. Soc. Am.* B, **1**, 366-389.

APPENDIX A: RELATIONSHIP BETWEEN YULE-FURRY AND LOGARITHMIC COUNTING STATISTICS

The statistical behavior of logarithmic counting statistics was shown in Sec. II A to resemble that of the pure-birth (Yule-Furry) model. The purpose of this appendix is to present a connection between the two counting distributions.

For a general discrete probability distribution $P(x)$, we consider the expectation

$$\left\langle \frac{x}{\langle x \rangle} \right\rangle = \sum_{x=0}^{\infty} \frac{x}{\langle x \rangle} P(x) = 1 , \tag{A1}$$

which is clearly normalized. This leads us to view the summed expression above as a probability transformation of the type

$$P_B(x) = [x/\langle x_A \rangle] P_A(x) , \tag{A2}$$

from which the auxiliary condition

$$\langle x_B \rangle = F_A + \langle x_A \rangle \tag{A3}$$

may be obtained, where F_A is the Fano factor, or variance-to-mean ratio, of the original distribution.

It is evident from the construction of Eq. (A2) that $P_B(x)$ is zero when $x = 0$, although this need not be true for $P_A(x)$. The former distribution may thus belong to the class of unit-shifted counting distributions, whose members can represent cascades in which the primary particle is included along with its offspring (as discussed in Sec. II A).

If $P_A(x)$ is the Poisson distribution, we employ Eqs. (1) and (3) to evaluate Eqs. (A2) and (A3):

$$P_B(x) = \langle x_A \rangle^{x-1} \exp(-\langle x_A \rangle)/(x-1)!, \quad x \geq 1 , \tag{A4a}$$

$$\langle x_B \rangle = 1 + \langle x_A \rangle . \tag{A4b}$$

This is just a unit-shifted version of the Poisson input, which can be easily verified by comparison with Eq. (5).

On the other hand, if $P_A(x)$ is the logarithmic distribution, then with the aid of many of Eqs. (7)–(13) we in fact obtain the Yule-Furry, or shifted Bose-Einstein, statistic:

$$P_B(x) = \exp(-\beta)[1 - \exp(-\beta)]^{x-1}, \quad x \geq 1 , \tag{A5a}$$

as long as

$$\langle x_B \rangle = 1 + \theta \langle x_A \rangle, \quad \text{or } \theta = \beta . \tag{A5b}$$

The simplicity of the transformation that links the Yule-Furry and logarithmic distributions is reflected by their shared mathematical properties.

We may recast Eq. (A2) into equivalent relations in the domain of generating functions. The most compact result is obtained via the probability-generating function (PGF), which is defined as

$$G(z) \equiv \langle z^x \rangle = \sum_{x=0}^{\infty} z^x P(x) . \tag{A6}$$

Applying Eq. (A6) to Eq. (A2) and using

$$G_B^{(s)}(z) = z G_B(z) , \tag{A7}$$

which relates the PGF's of the shifted and unshifted counting distributions, respectively, we obtain

$$G_B(z) = \frac{1}{\langle x_A \rangle} \frac{d}{dz} G_A(z) , \tag{A8a}$$

and its inverse

$$G_A(z) = \langle x_A \rangle \int_0^z G_B(z) dz . \tag{A8b}$$

From Eq. (A8b) we can see that the logarithmic PGF behaves like an unbounded cumulative version of the (unshifted) Bose-Einstein PGF. The moments of these two distributions can be linked to all orders by evaluating Eq. (A8) at $z = \exp(-s)$ to obtain relations for the moment-generating function $Q(s)$, whose properties are given by Eqs. (B1) and (B2) of Appendix B.

APPENDIX B: THIRD-ORDER MOMENTS FOR A CASCADED PROCESS

It is well known that the nth ordinary moment of a general counting distribution $P(x)$ may be obtained from its moment-generating function (MGF) (Ref. 14),

$$Q_x(s) \equiv \langle \exp(-sx) \rangle$$

$$= \sum_{x=0}^{\infty} \exp(-sx) P(x) , \tag{B1}$$

by successive differentiation according to

$$\langle x^n \rangle = (-1)^n \frac{\partial^n}{\partial s^n} Q_x(s) \Big|_{s=0} \quad \text{for all } n . \tag{B2}$$

Alternatively, it is oftentimes more convenient to employ the semi-invariant or cumulant-generating function (CGF)

$$K_x(s) \equiv \ln[Q_x(s)] , \tag{B3}$$

which yields the mean $\langle x \rangle$, variance $V(x)$, and third central moment $M_3(x)$ directly:

$$\langle x \rangle = -\frac{\partial}{\partial s} K_x(s) \Big|_{s=0} , \tag{B4a}$$

gradual decrease in probability between adjacent counts beyond $n = 1$ in all our experimental distributions; it is typically shallower than that required by the above estimate. We conclude that the afterpulsing contribution is not crucial and need not be considered in our models.

Luminescence can arise from the excitation and relaxation of molecular states in the faceplate material of the PMT, and has been studied extensively in connection with the effects of the radiation environment on spaceborne photomultiplier tubes.[34-37] Unlike Cherenkov photons, which are copiously produced but cannot be individually resolved at the detector, the less numerous luminescence photons are emitted during the comparatively long molecular relaxation times and may thus be registered as individual events. In fluorescence decay, this deexcitation commonly occurs from over a few to hundreds of microseconds, whereas in phosphorescence decay times may span from minutes to days. Of the two processes, the former is likely to be most significant to our counting experiments because its decay parameter can be well accommodated within the chosen counting times.

A statistical estimate of fluorescence noise is very difficult to obtain, unless the correlation between cosmic rays and fluorescence photons at the detector is known. This relationship cannot be deduced following a photon-counting experiment because knowledge of the occurrence times and pulse heights of individual events is lost. For this reason we cannot totally exclude fluorescence from our experimental measurements, although arguments against its dominance can be set forth.

If cosmic-ray particles were completely divorced from fluorescence emissions, the latter would simply add to the uniform background of random events, which is naturally incorporated into our models via the dark-rate parameter μ_D. A strong correlation between the two phenomena, on the other hand, would imply that a cosmic-ray count is closely followed by fluorescence events, thus falsely enhancing the shower clustering properties.

However, recent numerical simulations by Howell and Kennel of NASA (Ref. 35) have shown that existing correlations between Cherenkov and fluorescence photons in the PMT window are substantially reduced en route to detection at the photocathode. According to their simulated bivariate distribution, the latter may even at times appear without the presence of the former. "This," the authors suggest, "may be largely attributed to the directed nature of the Cherenkov radiation, in which photons exit the window with little or no internal reflection, but the fluorescence photons, in their more random flights, eventually strike the photocathode either directly or after reflection."

Some evidence in favor of this view may be gathered from our data. For instance, the experimental interevent time distributions shown in Figs. 14(b) and 15 reach local minima precisely within the temporal region associated with fluorescence photon arrivals (1–100 μsec). This is particularly true of the Hamamatsu tube, perhaps because its solar-blind response reduces fluorescence in the visible regime. Strong correlations would most cer-

tainly yield an enhancement in this emission-time window.

The small excess clustering that exists in the Apollo 17 counting distributions from space with very long counting times (see Fig. 2) also supports the above interpretation, especially since the MgF_2 window of the associated PMT fluoresces more efficiently than the SiO_2 and glass materials used by the other photomultipliers.

We believe fluorescence is not sufficient to fully account for the observed geometric feature that is so well described by our models (see Fig. 13). Nevertheless, future experiments employing point-process detectors in conjunction with pulse-height analyzers would more precisely elucidate the contribution of fluorescence noise to the clustered photomultiplier background.

IV. CONCLUSION AND DISCUSSION

We have shown that cosmic-ray-shower fluctuations at ground level can be effectively represented by the Poisson-driven Yule-Furry or the negative-binomial (Poisson-driven logarithmic) two-parameter counting probability distributions. The agreement between data and theory determined with photomultiplier tubes operated in the dark is remarkable, given the simplicity of the underlying pure-birth cascade model. The mathematical basis for the success of the theory lies in the presence of mean-square excess fluctuations in the variance of the multiplication statistics, of which the Yule-Furry and the logarithmic distributions are fundamental examples. In order to extract the cosmic-ray-shower component from the PMT data, we found it necessary to formulate a more precise description of photomultiplier background events by combining the Poisson-driven cascades with an independent Poisson random variable that summarizes other count sources at the detector.

Finally, we note that cosmic-ray clusters are particularly deleterious to some counting experiments. We have shown that cosmic-ray clusters can, in fact, be avoided in photon-counting experiments by restricting the counting duration to less than tens of seconds. The background events can then be described by a simple Poisson counting distribution with a very low mean (which is just a Bernoulli distribution when only 0 or 1 counts are observed). This is particularly useful for applications that seek to determine counting statistics which are sensitive to excess clustering. One important example is the detection of stationary photon-number squeezed light, as recently reported by a number of authors,[33,38-40] since the superposition of super-Poisson cosmic-ray counts at the phototube may destroy the nonclassical observable. In general, if photomultiplier tubes are used in photon-counting experiments that cannot be conducted at short durations, cosmic-ray clusters may pose a significant limitation. The background models we have presented will then provide a useful analytical point of departure.

Hamamatsu type R431-S PMT, as shown in Fig. 15.

A sharp peak appears in Fig. 14(b) at $\tau \approx 0.5$ μsec; this same phenomenon occurs, and is especially pronounced, in Fig. 15 at $\tau \approx 0.12$ μsec. This feature is attributed to brief but intense bursts of evenly spaced cosmic-ray events of unknown origin, which drive the Fano factor in counting experiments well above unity. These peaks permit us to estimate the typical intracluster event spacing. Since the smallest counting periods in our counting experiments $(T = 10$ μsec) are well in excess of the intracluster separation, it is clear that many of these events are indeed captured.

The broader tails of the two distributions (at the largest values of τ) are exponential and reflect intercluster separations arising from the cascade's primary (Poisson) process. Additional counts from noncascading sources at the PMT are also present, but these do not alter the exponential character of this tail since they are also random events. The quenching of both distributions below $\tau \approx 100$ nsec arises from the limited electronic resolution of the detectors. Typical pulse durations (full-width half

maximum) for the Hamamatsu system are rated at about 25 nsec.

E. Other sources of dark events in the photomultiplier tube

The phenomena of afterpulsing and luminescence, which are known to be important sources of spurious counts in certain photomultiplier tubes, have not been incorporated into our models. Since these are primarily tube-dependent effects, the introduction of further branching steps to the mathematical formulation may not always be necessary.

Afterpulsing is known to arise in photomultipliers from ion bombardment of the cathode and initial dynodes.[24,29] These charged agents result from the ionization of residual and leakage gases (such as hydrogen, helium, and nitrogen) during electromultiplication. The process is governed by the ion transit time, which is broadly distributed (even if afterpulsing from the dynodes is eliminated by proper discrimination) since it depends on the mass of the ionized species and its path length back to the photocathode.

A simple estimate of the afterpulsing probability $P_a(T)$ per photocathode event can be obtained by considering a cascade of Poisson primaries with a Bernoulli law for the creation of secondary afterpulses. Using a laser source of Poisson photons (to render the dark-rate contribution negligible) and conducting photon-counting experiments at $T = 1$ μsec and $D = 0.4$ sec (to enable cosmic-ray effects to be excluded from most of the experiments), Teich and Saleh[33] estimated that $P_a(1$ μsec$) \approx 1.6 \times 10^{-3}$ for the Hamamatsu tube. For larger counting times, an enhancement of $P_a(T)$ is expected, since more afterpulses are likely to be captured by the increased time window.

Even so, this is not sufficient to account for the very

(a)

(b)
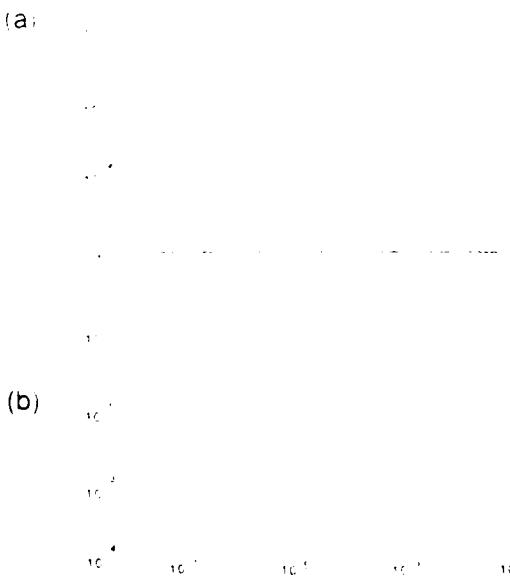
FIG. 14. (a) Interevent probability density $P(\tau)$ versus interevent time τ (sec) for background counts in a RCA type 8575 PMT, on semilogarithmic coordinates. The approximately exponential dependence for $\tau > 0.01$ sec reflects Poisson-distributed arrival times. The data set was collected on 7/19/82, at an 18:59:35 start time, and contain 881 002 samples. (b) The same data presented in a double-logarithmic format. The peak at $\tau \approx 0.5$ μsec reflects cosmic-ray intercluster event spacings. Even the smallest counting period used in our counting experiments $(T = 10$ μsec) is sufficiently large to capture these clustered events.
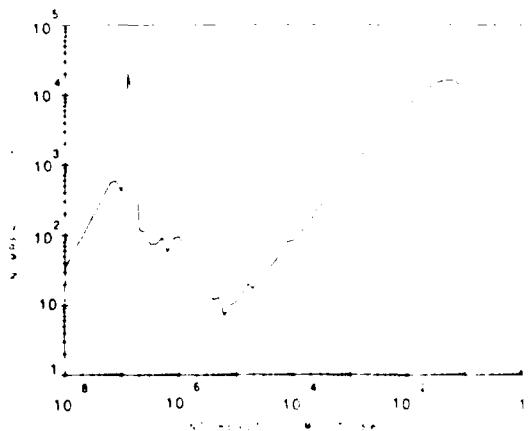
FIG. 15. Interevent probability density $P(\tau)$ versus interevent time τ (sec) for background counts in a Hamamatsu type R431-S PMT, on double-logarithmic coordinates. The distribution resembles that of Fig. 14(b). The pronounced feature arising from clustered cosmic-ray effects is evident at $\tau \approx 0.13$ μsec. The data set was collected on 9/7/83 and contains 500 000 samples.

(a)

(b)

(c)
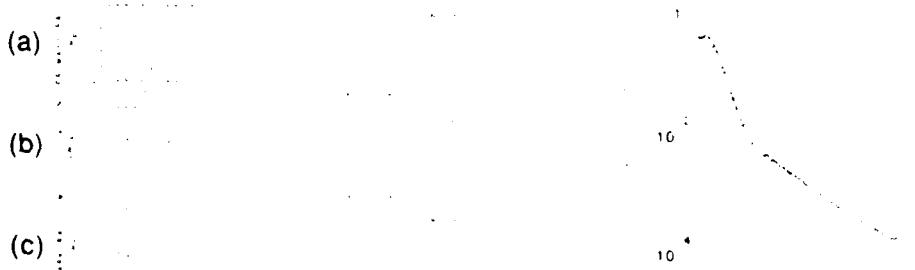
FIG. 11. (a) Primary cosmic-ray driving rate μ_p' (sec^{-1}), (b) mean multiplication $\langle m' \rangle$, and (c) dark (noncascading) driving rate μ_D' (sec^{-1}), versus counting period T. These parameters were extracted from the sequence of experiments shown in Fig 9 via the NB∘P model background events. They do not vary substantially over the range of counting times displayed. A total of six counting distributions out of 50 were not able to be fit by this model.

identification of the two components of each model. For low count numbers, the shape of the distribution is inherited primarily from the Poisson law of the noncascading events ($\langle n_D \rangle = \mu_D T \approx 1.2$, whereas $\langle n_p \rangle = \mu_p T \approx 0.054$). On the other hand, the quasigeometric tail beyond $n = 5$ is characteristic of the appropriate PDYF and negative-binomial distributions; it provides a telltale signature of cosmic-ray clusters at the detector. This dual behavior is quite adequately captured by the models we have considered.

We also remark on the manner of decomposition of the total background rate. In Table I it is shown that
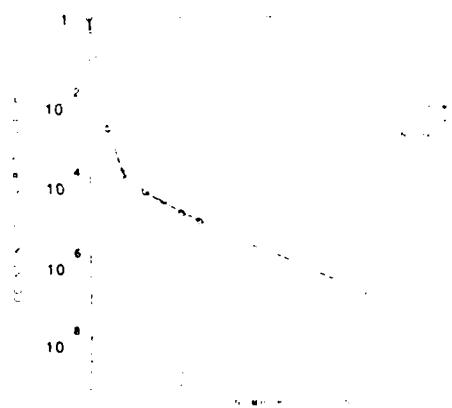
FIG. 12. Counting distribution $P(n;T)$ versus number of events n for the Schlumberger EMR type 541N-06-14 PMT designed for the NASA/JPL Galileo mission (9/16/77, Base line B-1 experiment) operated in the dark with counting period $T = 40$ μsec and duration $D = 19$ min. Open circles represent experimental values. Solid and dashed curves represent the PDYF∘P and NB∘P theoretical models, respectively. The experimental and theoretical parameters are displayed in Table I. This distribution resembles that for the Hamamatsu PMT displayed in Fig. 7 except that larger clusters are observed

FIG. 13. Counting distribution $P(n;T)$ versus number of events n for the RCA type 8575 PMT (7/8/82, 17:01:46) operated in the dark with counting time $T = 10$ msec and duration $D \approx 3$ d. Open circles represent experimental values. The solid and dashed curves represent the PDYF∘P and NB∘P theoretical models, respectively. All experimental and theoretical parameters are displayed in Table I. The underlying Poisson distribution for noncascading sources is clearly evident for low count numbers ($n \leq 5$), whereas the characteristic tail from the Poisson-driven cascade counting distributions is observed for large count numbers ($n \geq 5$). Both models capture this dual behavior with remarkable precision.

the dark rate for this counting distribution received the largest share in such a way as to maintain the primary rate estimate within an order of magnitude of those for the Hamamatsu tube experiments. This is not unreasonable since it is likely that noncascading sources (such as thermionic emission) and discriminator adjustments will vary from one experimental setup to another.

D. Interevent-time probability density functions

The experimental counting distribution $P(n;T)$ is an important statistical measure of a point process. The process of counting, however, forsakes a knowledge of the occurrence times of the individual events and, therefore, provides only a partial characterization of the entire point process. Another window on this information is the probability distribution $P(\tau)$ of the interevent times, which is constructed experimentally by recording a histogram for the temporal separations of adjacent pulses that are properly discriminated (see the discussion on pulse discrimination in Sec. III A). For the homogeneous Poisson point process, this distribution follows the simple exponential law $P(\tau) = \mu \exp(-\mu\tau)$, where μ is the constant driving rate of the process.[6]

In Fig. 14(a) we present $P(\tau)$ for the RCA type 8575 PMT operated in the dark. On these semilogarithmic coordinates, the experimental distribution may indeed be well approximated by a linear trend for large interevent times, but this Poisson description fails for $\tau \leq 0.01$ sec. The deviation can be scrutinized by a double-logarithmic presentation of this same data, as shown in Fig. 14(b). A similar distribution has been obtained with the

(a)

(b)
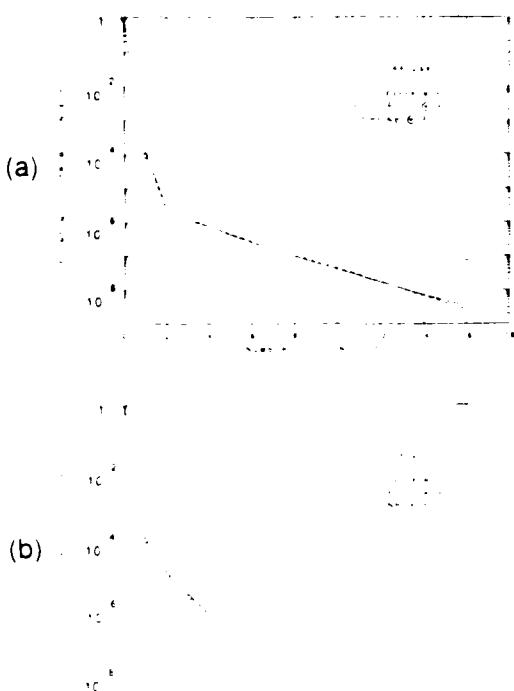
FIG 8 (a) Counting distribution $P(n;T)$ versus number of counts n for the Hamamatsu type R431-S PMT (11/14/83, 05 16 15) operated in the dark with counting time $T = 44$ μsec and duration $D = 1000$ sec. The open circles represent experimental values, whereas the solid and dashed curves represent the PDYF∘P and NB∘P theoretical models, respectively. All experimental and theoretical parameters are displayed in Table I. Note the presence of a dense cosmic-ray burst at $n = 16$ which reduces the quality of the fit. (b) Counting distribution for the same experiment after removal of the singular dense shower at $n = 16$ and renormalization of the data. Now note the similarity to Fig. 7. The Fano factor decreased from 1.54 to 1.38.

$\mu_p'(m') = 1.44$ sec^{-1} for the two models. These observed rates are an order of magnitude higher than that which is expected from an estimated sea-level flux of 1.0 cm^{-2} min^{-1} (Ref. 27) through the 2.5-cm Hamamatsu photocathode (i.e., 0.1 sec^{-1}). The discrepancy may arise from the superior resolution of this PMT operated in the photon-counting mode, which can detect low-height Cherenkov pulses.

5. Counting probability distributions for other PMT's

These observations are complemented by two additional counting distributions obtained with PMT's other than the Hamamatsu. The former used a counting period similar to that employed in our second sequence of experiments ($T = 40$ μsec, duration $D = 19$ min) with the Schlumberger EMR type 541N-06-14 PMT. This PMT is similar to that to be used in the now pending NASA/JPL Galileo mission to Jupiter. The distribu-

(a)

(b)
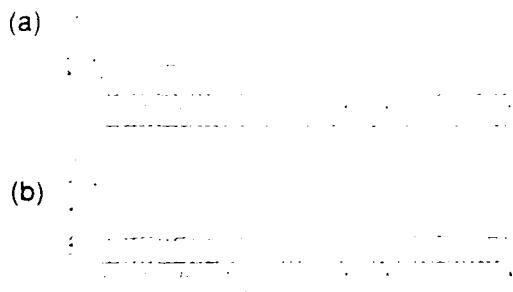
FIG. 9. (a) Mean count $\langle n \rangle$ (in units of 10^{-5}) and (b) Fano factor $F_n(T)$ versus counting time T (μsec) for the sequence of experiments illustrated in Fig. 6, after removal of large count numbers resulting from intense brief cosmic-ray bursts. The mean count was not significantly affected, but the occasional high peaks in the Fano factor that are prevalent in Fig. 6 are eliminated. Nevertheless, the counting distributions remain super-Poisson, with $F_n(T) \approx 1.3$, in the range of counting times displayed. The variation of T does not contribute to the raising of the Fano factor above unity; rather, this results from the increased experimental duration D.

tion, which is shown in Fig. 12, resembles that for the Hamamatsu PMT shown in Fig. 7, except that larger clusters are observed.

The final counting distribution, presented in Fig. 13, represents one of two similar experiments carried out with an RCA type 8575 tube. The counting time and duration were especially long ($T = 10$ msec, $D \approx 3$ d). Although the full distribution extended to the count number $n = 137$, it was curtailed at $n = 50$ and renormalized to increase clarity, but without loss of essential information.

This experiment is appealing because it facilitates an

(a)

(b)

(c)
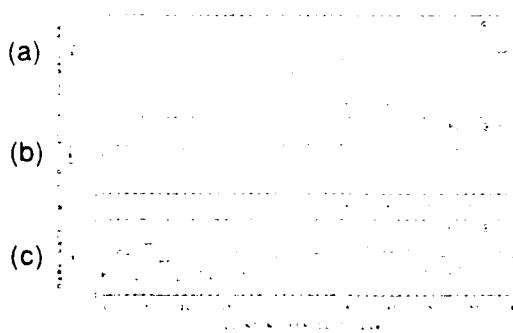
FIG. 10. (a) Primary cosmic-ray driving rate μ_p (sec^{-1}), (b) mean multiplication $\langle m \rangle$ and (c) dark (noncascading) driving rate μ_D (sec^{-1}), versus counting time T. These parameters were extracted from the sequence of experiments shown in Fig. 9 via the PDYF∘P model for background events. They do not vary substantially over the entire range of counting times displayed. Only one experiment (For $T = 24$ μsec) out of 50 was not able to be fit by the model.

(a)

(b)
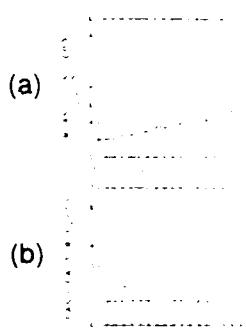
FIG. 6. (a) Mean number of events $\langle n \rangle$ (in units of 10^{-5}) versus counting time T (μsec) registered by a Hamamatsu type R431-S PMT (11/13/83, start time 19:30:00). The counting time T increases for each successive experiment beginning with $T = 10$ μsec and ending with $T = 59$ μsec, while the duration $D = 1000$ sec throughout. The time required to collect the entire set of 50 counting distributions was ≈ 14 h. The mean count has an approximately linear dependence on T with a slope of 3 sec^{-1}. (b) Fano factor $F_n(T)$ versus counting time T for the same set of experiments. The minimum value for the Fano factor is 1.19, whereas in Fig. 4(b) it is just the unity Poisson barrier. This is a consequence of the increase in duration to 1000 sec, which results in some cosmic-ray clusters always being present in each experiment, so that it is no longer possible to avoid their effect on the photomultiplier-tube background.

A typical counting distribution that lacked dense bursts is shown in Fig. 7 ($T = 41$ μsec; see Fig. 6). Both models share the ability to reproduce the nearly geometric tail of the distribution with excellent precision over 7 orders of magnitude. The presence of a dense burst, on the other hand, reduces the quality of both fits substantially, as the counting distribution of Fig. 8(a) illustrates ($T = 44$ μsec; see Fig. 6).

These large bursts do not, however, invalidate the theoretical approach. The cascade theory describes only the general properties of shower fluctuations and is not designed to accommodate the extremely nonstationary behavior of the dense cosmic-ray burst at $n = 16$ in this data set. Indeed, it is of interest to examine the data when a singular burst such as this is removed from the counting distribution. This is illustrated in the renormalized distribution shown in Fig. 8(b). The improvement is quite dramatic and the result resembles that presented in Fig. 7 (note the difference in scale). The mean and Fano factor that result from the removal of such bursts from the data of Fig. 6 is shown in Fig. 9. Comparing these figures shows that the removal of the dense cosmic-ray activity has little effect on the mean count but significantly reduces the Fano factor. Nevertheless, $F_n(T)$ retains its super-Poisson character, as a result of the residual tails of the counting distributions.

4. Extraction of model parameters

The uniformity of this revised data set encouraged us to attempt to extract the stationary properties of our
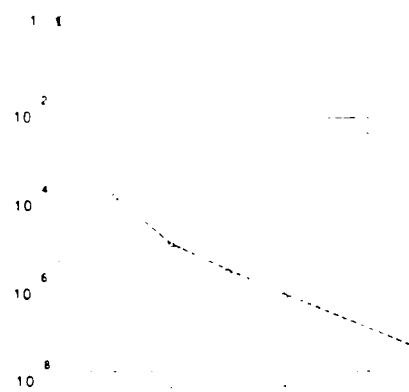
FIG. 7. Counting distribution $P(n;T)$ versus number of counts n for the Hamamatsu type R431-S PMT (11/14/83, 04:26:00) operated in the dark with counting time $T = 41$ μsec and duration $D = 1000$ sec. Open circles represent experimental values, whereas the solid and dashed curves represent the PDYF∘P and NB∘P theoretical models, respectively. All experimental and theoretical parameters are displayed in Table I. Both models provide excellent and nearly indistinguishable fits.

three model parameters. The total background-count rate is decomposed into a cosmic-ray primary rate and a dark rate from other sources at the detector. The last parameter carries the multiplicative strength of the atmospheric cascade. The theoretical parameters corresponding to the two models, $\{\mu_p, \langle m \rangle, \mu_D\}$ and $\{\mu_p', \langle m' \rangle, \mu_D'\}$, are given by Eqs. (32) and (35), respectively. These were extracted from each counting distribution represented in Fig. 9 and are displayed in Figs. 10(a)–10(c) and 11(a)–11(c) for the Poisson-driven Yule-Furry and negative-binomial models, respectively.

One immediate feature of the results is the stability of the mean multiplication that is obtained from both models over the entire range of counting times. From a mathematical point of view, the similarity in the magnitude of $\langle m \rangle$ and $\langle m' \rangle$ (both are ≈ 1.2) results from the proximity of the inversion parameter ($\Gamma \approx \frac{1}{4}$) to the region of convergence of the two solutions in Fig. 3.

Whatever small variation exists in the full background rate must, therefore, be apportioned to the primary and dark rates. Those for the PDYF ∘ P are about the same order of magnitude ($\mu_p, \mu_D \approx 1.0$ sec^{-1}), whereas the NB∘P model places a slightly greater emphasis on the primary rate ($\mu_p' \approx 1.2$ sec^{-1}) with a corresponding decrease in the dark rate ($\mu_D' \approx 0.7$ sec^{-1}). This comparative reduction in μ_D' has the effect of reducing the fitting efficiency of the NB∘P model, for its dark-rate estimate is more likely to be driven below zero on occasion, and thus yield an unacceptable solution. This is evident from the gaps in the histogram records of Figs. 10 and 11 (the PDYF∘P model successfully fit 49 of 50 experiments, while the NB∘P model could not accommodate five additional cases).

From the above parameters, we find the total fluctuation of cosmic-ray particles at ground level in this range of counting times to be about $\mu_p \langle m \rangle \approx 1.2$ sec^{-1} and

The resemblance between the Poisson-driven Yule-Furry and the negative-binomial descriptions of cosmic-ray showers is reflected by the near indistinguishability of the two fits in Fig. 5 and in those to follow. The convolution of the former with a Poisson distribution (PDY $\circ$ P) is shown by the solid curve, whereas the convolution with the latter (NB $\circ$ P) is shown by the dashed-line fit. We adopt this notation for all subsequent counting distributions. Parameters have been extracted from the data for each model in accordance with Eqs. (32) and (35); these are collected in Table I.

In a second photon-counting sequence, we increased the duration D of each experiment to 1000 sec and, furthermore, changed the counting time T from 10 μsec to 59 μsec in consecutive experiments, to observe the dependence of the model parameters on T. The number of samples per counting distribution naturally decreased with increasing counting time, but with an unsubstantial loss of accuracy; at $T = 59$ μsec we still obtained some 8 million samples as compared with 48 million samples at $T = 10$ μsec.

We present the mean count for this sequence of experiments in Fig. 6(a) as a function of counting time T. The data exhibit reasonable linearity between $\langle n \rangle$ and T, as expected; the larger the counting period, the more the

number of events from the point process that enter the counting window. The proportionality factor provides a measure of the mean rate of total background counts for this Hamamatsu tube. A linear fit to the data yields a slope of about 3.0 sec^{-1}, which accords with the estimate from Fig. 4(a).

The Fano factors in Fig. 6(b) can be compared to those of the first sequence of experiments in Fig. 4(b) (note the difference in the ordinate scales, however). The maximum value of $F_n(T)$ (6.98) occurred at $T = 52$ μsec (this resulted from single dense bursts of 62 and 64 counts each), whereas the minimum Fano factor was 1.19. It is evident that, under these experimental conditions, every measurement in the record leads to a super-Poisson result, so the effects of cosmic-ray showers on the photomultiplier-tube background can no longer be avoided. This is a direct consequence of the increase in the duration of each experiment to $D = 1000$ sec (≈ 17 min). It is longer than the entire sequence of experiments shown in Fig. 4(b), thereby unavoidably leading to the capture of cosmic-ray clusters.

The permanence of the cosmic-ray contribution throughout this data set provides an opportunity for testing the applicability of our models. We discuss two distinct examples for this tube.

TABLE I. Experimental parameters for the full counting distributions presented in Sec. III C.

Counting experiment	Fig	Counting time T	Duration of run D	Total samples N_s	Total rate $\to 1$ $\langle n \rangle /T$ (sec^{-1})	Fano factor $F_n(T)$	Gamma factor Γ
Hamamatsu R431-S 12/02/83 17:15:45	5	10 μsec	10 sec	480 353	2.498	1.500	0.333
Hamamatsu R431-S 11/14/83 04.26:00	7	41 μsec	1000 sec	11 732 575	2.732	1.355	0.310
Hamamatsu R431-S 11/14/83 05.16.15	8(a)	44 μsec	1000 sec	10 902 318	2.952	1.538	0.142
Hamamatsu R431-S 11/14/83 05:16:15 (burst deleted)	8(b)	44 μsec	1000 sec	10 902 317	2.918	1.373	0.341
EMR 541N-06-14 09/16/77	12	40 μsec	19 min	8 192 001	58.972	1.887	0.146
RCA 8575 07/08/82 17:01:46	13	10 msec	≈ 3 d	3 499 873	149.460	2.750	0.065

Counting experiment	Fig	Theoretical parameters extracted from fits to the data					
		μ_p (sec^{-1})	$\langle m \rangle$	μ_D (sec^{-1})	μ_F' (sec^{-1})	$\langle m' \rangle$	μ_D' (sec^{-1})
Hamamatsu R431-S 12/02/83 17:15:45	5	1.405	1.333	0.625	2.026	1.233	6.241×10^{-3}
Hamamatsu R431-S 11/14/83 04.26.00	7	0.844	1.408	1.544	1.236	1.282	1.148
Hamamatsu R431-S 11/14/83 05.16.15	8(a)	0.176	2.683	2.480	0.314	2.004	2.323
Hamamatsu R431-S 11/14/83 05.16.15 (burst deleted)	8(b)	1.338	1.310	1.166	1.919	1.218	0.582
EMR 541N-06-14 09/16/77	12	6.194	2.615	42.775	10.968	1.969	37.375
RCA 8575 07/08/82 17:01:46	13	5.406	5.444	120.030	11.990	3.273	110.220

ensure a reliable estimate of the distribution $P(n;T)$. The total time required to collect all samples is called the experimental duration D. In the case of measurements taken in the dark, of course, the PMT is not unshuttered.

We have also operated an RCA type 8575 PMT under similar conditions using a 2-kV Kepco type 188-0030 power supply, an EGG type NA201/N quadamplifier (8×), and an EGG type TR204/N discriminator. Additional data were obtained from a Schlumberger EMR type 541N-06-14 PMT being used in a study of radiation-environment effects on the star tracker for NASA's Galileo/Jupiter mission.[3]

The spectral response of all three tubes is within the Cherenkov line shape. The Hamamatsu device was selected for its high (Cs-Te) photocathode sensitivity in the ultraviolet 210 nm, whereas the other two PMT's employed bialkali materials that peak in the blue (400 nm).

3. Counting probability distributions

Two main sequences of experiments were carried out using the Hamamatsu tube. A discriminator setting of 1 V was used to ensure that the recorded events originate in the predynode environment. In the first sequence, the counting time T and the duration D of each experiment were fixed at 10 μsec and 10 sec, respectively. Given the duty cycle of the statistical counter (≈0.5) and a 5-sec processing hiatus between experiments, we were able to generate four counting distributions every minute with about 480 000 samples each. Collecting a full sequence in this set required over an hour, but we focus on the first 12 min, which are representative of the entire record.

Some of the experimental data is displayed (in histogram form) as a function of elapsed time in Fig. 4. We depict the mean count $\langle n \rangle$ of each counting distribution $P(n;T)$ in Fig. 4(a) and the associated Fano factor $F_n(T)$ in Fig. 4(b), both with a logarithmic ordinate. The total experimental background rate is ≈2 sec^{-1}.

The Fano factor serves as an effective tool for discerning the underlying clustering phenomena. Although it is essentially unity in nearly 70% of these experiments, the Fano factors for the remaining counting distributions are sufficiently greater than unity to signal the presence of shower events. An example is provided about halfway into the record, where there was an obvious deviation from Poisson behavior that lasted for over a minute. The shower was most intense during the counting experiment that began at the fifth minute; in a single counting time of 10 μsec there was a burst of 32 events and the Fano factor reached a value of 21. Our choice of experimental parameters is clearly sufficient to resolve the characteristic arrivals of shower particles at the detector. This behavior could not have been inferred by analyzing the evolution of the mean count alone, as is evident from the record in Fig. 4(a).

Some of the counting distributions are, however, too sparse to provide a rigorous test of the models under consideration.[32] In fact, only one super-Poisson experiment in the 12-min record was successfully fit by our models; it is shown in Fig. 5 and corresponds to 5:45 elapsed time in Fig. 4. Although a triple count occurred only once in that experiment, it was sufficient to drive the Fano factor above unity and exemplifies the vulnerability of Poisson statistics to excess clustering.

(a)

(b)

FIG. 4. (a) Mean number of events $\langle n \rangle$ (in units of 10^{-5}) versus real time t (min) registered by a Hamamatsu type R431-S PMT over a 12-min period (12/2/83, start time 17:10:00). The counting time $T = 10$ μsec and the duration $D = 10$ sec for each experiment. Successive experiments take place every 15 sec. The total background rate is ≈2 sec^{-1}. (b) Fano factor $F_n(T)$ versus real time for the same set of experiments. Note that $F_n(T) = 1.0$ for about 70% of the counting distributions, indicating the complete absence of cosmic-ray clusters. It stretches to a maximum of 21 for the remaining runs, indicating the presence of bursts. The cosmic-ray contribution to the photomultiplier background can thus be generally avoided if the duration of an experiment is limited to less than tens of seconds.
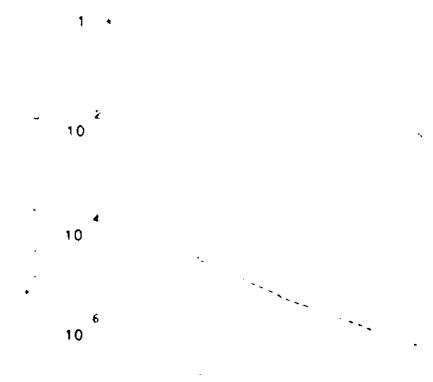
FIG. 5. Counting distribution $P(n;T)$ versus number of counts n for the Hamamatsu type R431-S PMT (12/2/83, 17:15:45) operated in the dark with count time $T = 10$ μsec and duration $D = 10$ sec. Open circles represent experimental values (note that $P(2) = 0$); the convolution of the Poisson-driven Yule-Furry cosmic-ray cascade with a Poisson (PDYF∘P) is shown as a solid curve, whereas the negative-binomial cascade convolved with a Poisson (NB∘P) leads to the dashed curve. All experimental and theoretical parameters are displayed in Table I. The single triple count cluster was sufficient to drive the Fano factor to 1.5.

The inversion is facilitated by defining a convenient auxiliary parameter

$$\Gamma = [V(n) - \langle n \rangle] / [M_3(n) - V(n)] , \quad (30)$$

which is readily calculated from the statistics of $P(n;T)$. Assuming first a Poisson-driven Yule-Furry cascade, we substitute Eq. (23) into Eq. (29) to obtain

$$\langle n \rangle = \mu_D T + \mu_f T \exp(\beta) , \quad (31a)$$

$$V(n) = \mu_D T + \mu_f T \exp(\beta)[2 \exp(\beta) - 1] , \quad (31b)$$

$$M_3(n) = \mu_D T + \mu_f T \exp(\beta)[6 \exp(2\beta) - 6 \exp(\beta) + 1] . \quad (31c)$$

Combining the above via Eq. (30) provides a quadratic equation for the mean multiplication $\langle m \rangle = \exp(\beta)$ with the simple roots

$$\langle m \rangle = 1 , \quad (32a)$$

$$\langle m \rangle = \tfrac{1}{2}[1 + (1/\Gamma)] . \quad (32b)$$

Equation (32a) represents the absence of branching and is uninteresting; Eq. (32b) is, therefore, the unique solution for the nontrivial multiplicative nature of the cascade as seen by the PMT. The associated driving rates are then determined from

$$\mu_f T = [(3\Gamma)^2/2(1 + \Gamma)(1 - 2\Gamma)](F_n - 1)\langle n \rangle , \quad (32c)$$

$$\mu_D T = \{[(2 - \Gamma) - (3\Gamma F_n)]/[2(1 - 2\Gamma)]\}\langle n \rangle , \quad (32d)$$

where F_n is the experimental Fano factor. Repeating the technique for the negative-binomial cascade, we employ Eq. (26) to yield

$$\langle n \rangle = \mu_D T + \langle n_0 \rangle , \quad (33a)$$

$$V(n) = \mu_D T + \langle n_0 \rangle[1 + (\langle n_0 \rangle / M)] , \quad (33b)$$

$$M_3(n) = \mu_D T + \langle n_0 \rangle[1 + (\langle n_0 \rangle / M)][1 + 2(\langle n_0 \rangle / M)] . \quad (33c)$$

Solving for the negative-binomial parameters, we obtain

$$\langle n_0 \rangle = [(2\Gamma)/(1 - 2\Gamma)](F_n - 1)\langle n \rangle , \quad (34a)$$

$$M = [(2\Gamma)/(1 - 2\Gamma)]^2(F_n - 1)\langle n \rangle . \quad (34b)$$

Returning to the parameter of the logarithmic distribution via Eq. (25) then yields $\theta = -\ln(2\Gamma)$, so that

$$\langle m' \rangle = [\ln(2\Gamma)]^{-1}[1 - (1/2\Gamma)] , \quad (35a)$$

$$\mu_f' T = \ln[1/(2\Gamma)][(2\Gamma)/(1 - 2\Gamma)]^2 F_n - 1)\langle n \rangle , \quad (35b)$$

$$\mu_D' T = \{[1 - (2\Gamma)F_n]/(1 - 2\Gamma)\}\langle n \rangle \quad (35c)$$

The resemblance between Eqs. (32) and (35) is a direct consequence of the similar underlying multiplication models that are discussed in Sec. II A. We have primed the latter three parameters to minimize confusion between the two models. Graphical representations for the mean multiplications $\langle m \rangle$ and $\langle m' \rangle$ are presented in

Fig. 3, as a function of Γ, for comparison. The region of physical significance is shaded and bounded on the left by infinite multiplication for $\Gamma = 0$ and on the right by unity multiplication for $\Gamma = \tfrac{1}{2}$. This follows clearly from the expressions $\Gamma = 1/[3 \exp(\beta) - 1]$ and $\Gamma = \tfrac{1}{2}\exp(-\theta)$ when $0 < \beta < \infty$ and $0 < \theta < \infty$, respectively. The logarithmic factor in the negative-binomial result modulates the solution in just the right way to produce a graphical trend like that for the PDYF in this domain.

2. Experimental apparatus

We now present the results of photon-counting experiments conducted with several different photomultipliers in the dark at ground level. A Hamamatsu type R431-S tube was operated at ambient temperature on the 13th floor of the Seeley W. Mudd building of Columbia University. It was inserted in a Hamamatsu type C1050 photon-counting base that provides preamplification, discrimination, and pulse shaping. An Ortec type 456 power supply delivered 1025 V to the PMT. The output of the PMT base was passed through a buffer amplifier $(5\times)$ and fed to a Hewlett-Packard type 5370A rate counter and a Langley-Ford type 1096 statistical photon counter. The former measures the detected mean photon count rate, whereas the latter generates the probability distribution $P(n;T)$. The experiment was controlled by a Hewlett-Packard type 9825B minicomputer.

Photon counting is a simple experimental procedure. Each sample of the counting distribution is obtained by unshuttering the photosensitive PMT cathode and recording the number of appropriately discriminated events that occur within a fixed counting time T. By repeating the procedure, enough samples are then taken to
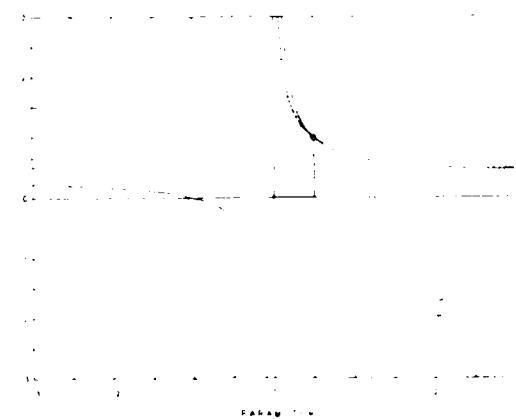
FIG. 3. Graphical dependence of the mean multiplication solutions $\langle m \rangle$ and $\langle m' \rangle$ on the experimental inversion parameter Γ. The solution represented by the solid curve is obtained if the Poisson-driven Yule-Furry (PDYF) cascade is incorporated into a model for photomultiplier background events, whereas the solution represented by the dashed curve results if the negative-binomial (NB) cascade is considered. The region of physical significance, $\tfrac{1}{2} > \Gamma > 0$, is shaded and bounded by unity and infinite mean multiplications respectively. The two solutions exhibit similar behavior in this region.

with spacecraft altitude and attitude, but was independent of the position of the spacecraft in lunar orbit; that is, the background signal was clearly not of solar, lunar, or earth origin." The count rate varied from 20.7 to 28.3 sec^{-1} at lunar altitudes of 26 and 312.5 km, respectively, as compared with 0.6 sec^{-1} in prelaunch laboratory tests.

The PMT background events were processed into counting probability distributions with a counting time $T = 0.1$ sec, two of which are reproduced in Fig. 2. The dashed line, which represents a Poisson distribution with the same mean as the experimental data, provides a good fit in both cases. A small excess clustering is inferred from the raised tail of the experimental distributions and their associated Fano factors, which are slightly greater than unity [1.08 and 1.13 for Figs. 2(a) and 2(b), respectively]. Fastie attributes this super-Poisson effect to additive fluorescence photons from the MgF_2 window of the photomultiplier tube. We reserve a discussion of fluorescence for a later section of this paper and accept Fastie's data as confirmation of the independence of primary cosmic-ray arrivals.

C. Cosmic-ray counting experiments at ground level

1. Models for noise in photomultiplier tubes operated in the dark

Photomultiplier-tube background events are most often assumed to obey a simple homogeneous Poisson law, with dark-count index n_D, reflecting the presence of independently arriving events generated by thermionic emission at the photocathode and dynode stages, by radioactive elements such as ^{40}K in the window material, and by a number of other effects, some of which are discussed subsequently.[24,25,27-29] The counting distribution $P(n_D; T)$ then takes the form of Eq. (1).

However, the data at ground level do not support this simple model. Rather, it is necessary to extend this framework by incorporating cosmic-ray counts from a Poisson-driven cascade. The generalization proceeds in a twofold fashion. The cascading component of the cosmic rays has been shown to exhibit super-Poisson behavior via a counting distribution $P(n_c; T)$ that incorporates the clustered aspect of the background. A noncascading cosmic-ray component, such as the muon, is also registered at the PMT. These events are essentially independent and random and may therefore be fully absorbed in the Poisson counting distribution $P(n_D; T)$ used to describe the thermionic emission.

The generalization is provided by the independent addition of the dark index n_D with the cascade index n_c. The overall counting distribution $P(n; T)$ at ground level is then the convolution of these distributions

$$P(n; T) = P(n_c; T) \circledast P(n_D; T) . \tag{28}$$

Their statistics up to third order are purely additive,[30] i.e.,

$$\langle n \rangle = \mu_D T + \langle n_c \rangle , \tag{29a}$$
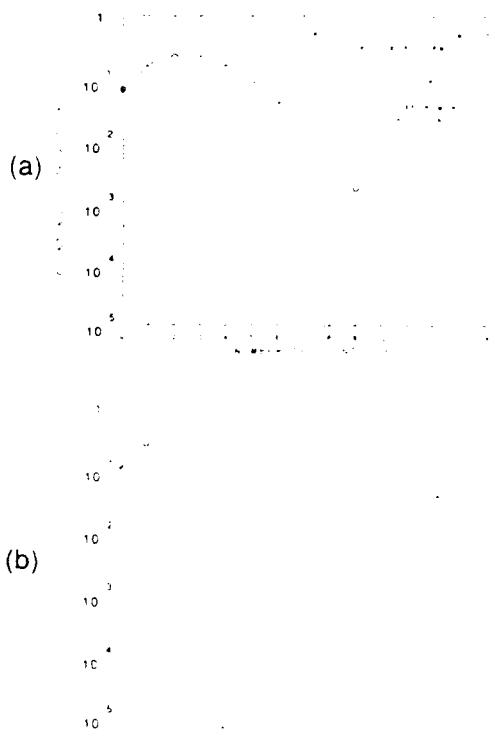
FIG. 2. (a) Sample 1 counting distribution $P(n; T)$ versus number of counts n from a Schlumberger EMR type 542G-09 photomultiplier tube in space, operated with counting time $T = 0.1$ sec (data adapted from Table II of Ref. 28. The open circles represent experimental values from 5750 total samples. The mean count $\langle n \rangle = 2.85$ and the Fano factor $F_n(T) = 1.08$. The dashed curve is a Poisson distribution with the same mean; it represents an excellent fit to the data and confirms the independence of primary cosmic-ray arrivals. (b) Sample 2 counting distribution from the same source. In this case, the experimental mean count $\langle n \rangle = 2.32$ and the Fano factor $F_n(T) = 1.13$ from 5750 total samples. Fastie suggests that the excess clusters represented by the raised tail of the distribution may be caused by luminescence in the tube window.

$$V(n) = \mu_D T + V(n_c) , \tag{29b}$$

$$M_3(n) = \mu_D T + M_3(n_c) . \tag{29c}$$

The analytical form of $P(n; T)$ will clearly depend on the choice of the Poisson-driven cascade model.

Given the experimental moments obtained from the counting distribution $P(n; T)$ of a PMT operated in the dark, it is of interest to invert the relations in Eq. (29) to obtain a set of model-dependent parameters μ_r (the primary driving rate), μ_D (the dark driving rate arising from noncascading sources), and $\langle m \rangle$ (the mean multiplication per primary). The large spatial extent of the shower and the small solid angle of acceptance of the detector lead us to expect that the parameters μ_r and $\langle m \rangle$ will not reflect the full effects of the atmospheric cascade process. Nevertheless, sufficient cosmic-ray clustering remains to control these parameters.

$$M_3(n_c) = \mu_p T \exp(\beta)[6\exp(2\beta) - 6\exp(\beta) + 1] .$$

$$(23c)$$

These formulas are appropriate for the Poisson-driven Yule-Furry (PDYF) counting distribution, which has been shown to obey the recursion relation[18,19]

$$P(n_c+1;T) = \frac{\langle n_c \rangle}{n_c+1} \sum_{k=0}^{n_c} P(n_c-k;T) A(k) , \quad (24a)$$

$$A(k) = (k+1)\exp(-2\beta)[1-\exp(-\beta)]^k , \quad (24b)$$

with the initial condition

$$P(0;T) = \exp(-\mu_p T) . \quad (24c)$$

Equation (24c) indicates that the probability of observing zero cascade counts is simply the probability of arrival of zero primaries in the counting interval. Note also that as $\beta \to 0$ this recursion relation reduces to that for the Poisson law of the primaries.

Finally, we consider the case of logarithmic multiplication with statistics given by Eq. (12). Utilizing Eq. (20) once more and using the transformation

$$\langle n_0 \rangle = \mu_p T[\exp(\theta)-1]/\theta , \quad (25a)$$

$$M = \mu_p T/\theta , \quad (25b)$$

we find

$$\langle n_c \rangle = \langle n_0 \rangle , \quad (26a)$$

$$V(n_c) = \langle n_0 \rangle [1 + \langle n_0 \rangle /M] , \quad (26b)$$

$$M_3(n_c) = \langle n_0 \rangle [1 + \langle n_0 \rangle /M][1 + 2\langle n_0 \rangle /M] , \quad (26c)$$

to be the representative statistics of the Poisson-driven logarithmic distribution. This turns out to be none other than the negative-binomial counting distribution,[14] with mean $\langle n_0 \rangle$ and degrees-of-freedom parameter M:

$$P(n_c;T)$$
$$= \frac{(n_c+M-1)!}{n_c!(M-1)!} \left[1 + \frac{\langle n_0 \rangle}{M} \right]^{-M} \left[1 + \frac{M}{\langle n_0 \rangle} \right]^{-n_c} ,$$

$$(27)$$

as first shown by Quenouille.[20] Interestingly enough, the negative-binomial distribution also arises, in the context of cascade theory, as the steady-state limit of the problem of birth, death, and immigration with stationary rates.[21] The result is independent of the number of primary particles and is only valid if the death rate exceeds the birth rate; the limit thus represents a balance between the decaying cascade population and the immigration rate. The Poisson-driven Yule-Furry and the negative-binomial distributions are both two-parameter distributions.

III. EXPERIMENTS

A. The photomultiplier tube as a detector of cosmic rays

It has long been known that cosmic-ray-shower particles at ground level are registered by optical detectors

such as the photomultiplier tube (PMT). The underlying reason is understood from the work of Cherenkov in 1934. A relativistic charged particle which travels at a speed exceeding the phase velocity of light in a transparent medium can induce a broadband optical shock wave with a spectral line shape predominant in the visible and ultraviolet regions of the electromagnetic spectrum.[22,23] Since this Cherenkov emission is highly directive and efficient, a single charged energetic shower component that traverses the thin photomultiplier-tube window generates a dense packet of photons that is subsequently amplified into a large anode current pulse.

Because the photomultiplier cannot resolve the individual photons from the brief Cherenkov flash, which lasts of the order of tens of picoseconds in accordance with the cosmic ray's transit time, the output signal is just the device's impulse response with an amplitude proportional to the number of Cherenkov-induced photoelectrons (if saturation effects are not considered). These large pulses in fact account for the pronounced tails in pulse-height distributions of the photomultiplier background.[24-26]

Young[27] has shown that the output current from a PMT, operated in the continuous mode in the dark, is indeed marked by substantial fluctuations that can seriously degrade the signal-to-noise ratio in a variety of experiments. As a photon counter, however, the PMT offers a significant reduction in the Cherenkov noise, since each output pulse that exceeds an adjustable threshold height yields a single count, *regardless of its excess amplitude*. The passage of a cosmic ray through the photon-counting PMT therefore registers one, *and only one*, count, although these background events may still be highly detrimental to some counting and correlation experiments. Barring other sources of PMT events, a counting distribution from the device in the dark thus provides a direct probe of the local cosmic-ray activity, and, in this capacity, the photon counter is the ideal tool for extracting its clustering properties.

Previous efforts by Young and others to characterize cosmic-ray effects at the photomultiplier tube were intended primarily for dc and mean-square current applications that do not take full advantage of pulse-height discrimination. The theoretical counting distributions considered in the previous section should provide a more appropriate and complete description of the cosmic-ray noise, since they capture the essential clustering behavior of these particles at ground level and retain their counting statistics to all orders.

B. Cosmic-ray counting experiments in space

Before examining experimental configurations at ground level in which Poisson-driven cascades play a role, it is useful to experimentally verify that the primary cosmic-ray flux above the atmosphere does indeed obey the Poisson law.

Fastie[28] conducted dark-counting experiments in space using a Schlumberger EMR type 542G-09 solar-blind PMT from the ultraviolet spectrometer aboard the Apollo 17 space-vehicle command module. He concluded that "... in lunar orbit the observed dark count varied

$$\langle m \rangle = I(t_s) , \tag{15a}$$

$$V(m) = \langle m \rangle + \langle m \rangle^2 \left[2 \int_0^{t_s} \frac{\lambda(t)dt}{I(t)} - 1 \right] , \tag{15b}$$

where

$$I(t) = \exp \left[\int_0^t [\lambda(t') - v(t')]dt' \right] , \tag{15c}$$

where $\lambda(t)$ and $v(t)$ are the birth and death rates, respectively. Setting $v(t) = 0$ and $\lambda(t) = \lambda$ reproduces the Yule-Furry results presented in Eqs. (8a) and (8b).

The temporal variation of these rates can be appropriately chosen to represent the degradation in offspring production due to the partitioning of energy among the particles of a developing cascade. Arley considered a simple example in which the death rate is linearly infused [$v(t) = vt$] while the birth rate is held constant [$\lambda(t) = \lambda$]. The evolution of the mean multiplication can be solved using Eq. (15a)

$$\langle m \rangle = \exp(\lambda t_s) \exp(-vt_s^2/2) . \tag{16}$$

This is the same result as that for a birth process that attenuates via the tail of a zero-mean Gaussian envelope with the standard deviation $\sigma = \sqrt{1/v}$. Since more energetic primaries yield larger showers, the Gaussian modulation can further be made dependent on the total primary energy by proper adjustment of the death-rate scale.

Apparently the simple geometric character of the Yule-Furry pure-birth solution offers the essential ingredient for the characterization of cosmic-ray showers, regardless of the level of sophistication incorporated into the cascade model. Arley appreciated this and suggested that "... the region of validity of the simplifying approximations, of which any theory must necessarily make use, is always far wider than might be justified by theoretical arguments."

Mathematically, cascade counting distributions support a multiplication variance term that is proportional to the square of the mean. This is evident in Eq. (15b) for arbitrary birth and death rates. In the hierarchy of discrete probability distributions, the Bose-Einstein distribution (as well as its shifted version, the Yule-Furry) provides the simplest example of mean-square excess fluctuations. Our intuition regarding the logarithmic description of the multiplication is based on the presence of a similar dependence, as can be inferred from Eq. (12b).

B. Showers initiated by a Poisson number of primaries

We now generalize the results of the previous section by considering a number of primary events p, which is randomized via a suitable counting distribution $P(p;T)$. Since each primary event independently multiplies according to a model counting distribution $P(m;T)$, the overall cascade is an example of a compound process with well-known properties. In particular, the moment-generating function $Q_c(s)$ of the overall cascade counting distribution can be obtained by nesting those of the primary and multiplication counting distributions according to

$$Q_c(s) = Q_p(-\ln Q_m(s)) , \tag{17}$$

where the moment-generating function (MGF) is defined as

$$Q_x(s) \equiv \langle \exp(-sx) \rangle$$
$$= \sum_{x=0}^{\infty} \exp(-sx)P(x;T) , \tag{18}$$

for a general counting process x (Ref. 14).

We can use the above result to obtain a useful set of equations that expresses the mean, variance, and third central moment of the overall cascade counting distributions $P(n_c;T)$ in terms of those of the constituent counting processes:

$$\langle n_c \rangle = \langle p \rangle \langle m \rangle , \tag{19a}$$

$$V(n_c) = V(p)\langle m \rangle^2 + \langle p \rangle V(m) , \tag{19b}$$

$$M_3(n_c) = M_3(p)\langle m \rangle^3 + 3V(p)\langle m \rangle V(m) + \langle p \rangle M_3(m) . \tag{19c}$$

These results are quite general and are explicitly derived in Appendix B. The former two expressions were first proved by Shockley and Pierce,[15] and the second is known as the cascade variance theorem.[16] Equation (19c) is an extension to third-order statistics; higher-order theorems are discussed via an equivalent method by Shih[17] in his treatment of hadronic multiplicity distributions.

If we now specify that the primaries arrive in *stationary* (or homogeneous) Poisson form, Eq. (19) simplifies considerably. We denote the mean of the primary Poisson distribution $\langle p \rangle$ as $\mu_p T$, where generally μ is the rate parameter of the homogeneous Poisson point process (HPP) and T is the counting time over which we observe the events. For the Poisson primaries, we then have $\langle p \rangle = V(p) = M_3(p) = \mu_p T$, whereupon

$$\langle n_c \rangle = \mu_p T \langle m \rangle , \tag{20a}$$

$$V(n_c) = \mu_p T[\langle m \rangle^2 + V(m)] , \tag{20b}$$

$$M_3(n_c) = \mu_p T[\langle m \rangle^3 + 3\langle m \rangle V(m) + M_3(m)] . \tag{20c}$$

This can be compactly rearranged in terms of the ordinary moments as

$$\langle n_c \rangle = \mu_p T \langle m \rangle , \tag{21a}$$

$$V(n_c) = \mu_p T \langle m^2 \rangle , \tag{21b}$$

$$M_3(n_c) = \mu_p T \langle m^3 \rangle . \tag{21c}$$

The Fano factor is given by

$$F_c(T) = \langle m \rangle + F_m(T) , \tag{22}$$

and depends solely on the multiplication statistics.

For the case of a primary Poisson counting distribution with Yule-Furry multiplication, we use Eq. (8) in Eq. (20) to provide the overall cascade statistics, up to third order:

$$\langle n_c \rangle = \mu_p T \exp(\beta) , \tag{23a}$$

$$V(n_c) = \mu_p T \exp(\beta)[2\exp(\beta) - 1] , \tag{23b}$$

$$P(m;T) = (1/\theta)(q^m/m), \quad 0 \leq q \leq 1, \, m \geq 1, \quad (10a)$$

$$\theta = \ln[1/(1-q)], \quad (10b)$$

where θ is a normalization factor. We can rewrite this distribution, in terms of θ, into the suggestive form

$$P(m;T) = [1 - \exp(-\theta)]^m/\theta m, \quad \theta \geq 0, \, m \geq 1, \quad (11)$$

with statistics, up to third order,

$$\langle m \rangle = (1/\theta)[\exp(\theta) - 1], \quad (12a)$$

$$V(m) = \langle m \rangle + (\theta - 1)\langle m \rangle^2, \quad (12b)$$

$$M_3(m) = \langle m \rangle + 3(\theta - 1)\langle m \rangle^2 + (2\theta^2 - 3\theta + 2)\langle m \rangle^3. \quad (12c)$$

The corresponding Fano factor is

$$F_m(T) = 1 + [(\theta - 1)/\theta][\exp(\theta) - 1]. \quad (13)$$

The striking similarity between the Yule-Furry and logarithmic statistics is not accidental. As shown in Appendix A, these counting distributions are in fact related when $\theta = \beta$, although statistical equality does not follow from this condition [as is evident from Eqs. (8a) and (12a)]. Nevertheless, we may mimic the evolution of the Yule-Furry mean multiplication by setting

$$\langle m \rangle = (1/\theta)[\exp(\theta) - 1] = \exp(\gamma), \quad (14)$$

which is a transcendental equation for $\theta = f(\gamma)$. The dependence of the Fano factor on γ cannot be determined analytically, although this poses no difficulty since for a given value of γ we may obtain the corresponding value of θ (and, hence, the Fano factor) from a numerical table.

The advantage of this technique rests on the mathematical equivalence of the parameter γ to α and β, which allows us to present the associated Fano-factor dependencies together for comparison in Fig. 1(a). In contrast with the asymptotic Fano factor of the shifted-Poisson counting distribution, the Fano factors for the Yule-Furry and logarithmic counting distributions follow a similar trend and rise quickly with increasing multiplication. The latter two models thus allow us to describe the clustered behavior of cosmic-ray showers for large multiplication with far greater accuracy than the former. The principal underlying reason is the association of parents to their offspring at each branching stage of the multiplicative process.

Representative probability distributions are presented in Fig. 1(b) for $\langle m \rangle = 1.005$, 1.2, and 3.0. Note that in this, and all counting distributions to follow, the theoretical values are shown as continuous curves for clarity, but they have significance only at the discrete values of the count index. It is clear that for fixed multiplication the logarithmic distribution exhibits greater variance than the Yule-Furry distribution; however, all three models converge to an equivalent description for very low multiplication. This is also apparent in Fig. 1(a).

We conclude this section by speculating on the validity of the Yule-Furry distribution in describing the soft component of the cosmic-ray progeny. Clearly the mod-
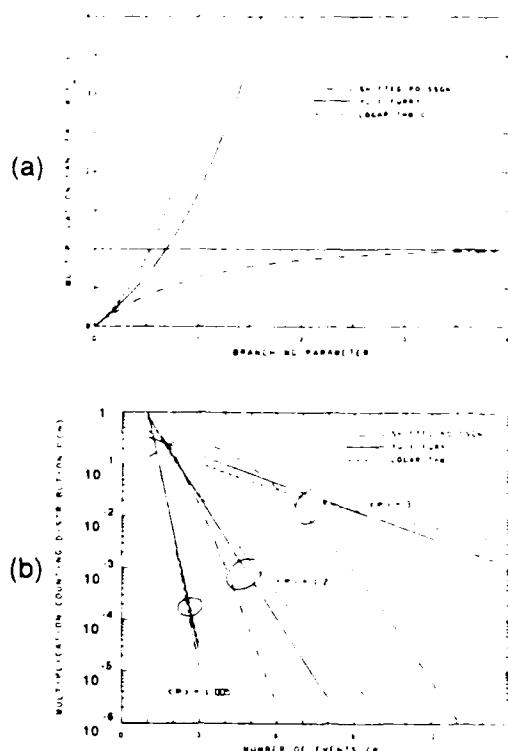

FIG. 1. (a) Multiplication Fano factor $F_m(T) \equiv V(m)/\langle m \rangle$ versus branching parameter for three models of a cascade that is initiated by a single primary particle. The Fano factors of the Yule-Furry (solid curve) and logarithmic (dashed curve) counting distributions increase rapidly with branching parameter, in contrast with the bounded Fano factor of the shifted-Poisson (dashed-dotted curve) counting distribution. The former two models thus provide a more accurate description of cosmic-ray shower fluctuations. (b) Representative multiplication counting distributions $P(m;T)$ versus number of multiplication events m for the models of (a) with $\langle m \rangle = 1.005$, 1.2, and 3.0. Although all three models are equivalent at low mean multiplication, the Yule-Furry and logarithmic differ substantially from the shifted-Poisson counting distribution as $\langle m \rangle$ increases. The tails of distributions for the former two models become much more developed than the latter, thus incorporating cascades with a larger number of events.

el represents only a first-order idealization of the problem since it considers only a single particle species (hence, not accounting for the alternating generations of electrons and photons in the physical cascade), and, as a pure-birth process, it does not incorporate particle losses and immigration.

Extensions of the Yule-Furry have been considered by Arley[11] and Srinivasan[12] in their authoritative monographs. The framework for more sophisticated cascade treatments can be constructed from the general problem of birth, death, and immigration with time-dependent rates.[13] When no immigration is allowed, the mean and variance of the multiplication can be written more succinctly, yet with no loss of generality, as

that a primary cosmic ray collisionally excites only a single shower.

We employ the Poisson-driven Yule-Furry and negative-binomial counting probability distributions, which are derived in the context of cascade theory in the next section, to describe the fluctuations of cosmic-ray showers at ground level. To demonstrate the efficacy of these models, we present in Sec. III experimental counting distributions, which were generated in the dark from ordinary photomultiplier tubes, over a broad range of counting times and experimental durations.

II. THEORY OF COSMIC-RAY-SHOWER STATISTICS

A. Showers initiated by a single primary event

The first statistical treatment of a single-primary-event cascade as applied to cosmic rays was presented by Bhabha and Heitler[4] in the United Kingdom and simultaneously by Carlson and Oppenheimer[5] in the United States. The model is equivalent to a zero-memory process whereby daughter particles are independently added to the cascade. In a counting experiment, this model leads to a Poisson-distributed number of secondaries.[6,7] i.e.,

$$P(k;T) = \langle k \rangle^k \exp(-\langle k \rangle)/k!, \quad k \geq 0 . \quad (1)$$

The angular brackets denote an ensemble average and T is the counting time, which is chosen to be sufficiently long to capture the entire cluster of secondary events, so their nonstationary production can be effectively treated as instantaneous. In other words, if t_s is the temporal spread of the cluster, then the requirement $T \gg t_s$ ensures that no events are lost to the counting process. We assume this to hold henceforth.

Clearly the conservation laws that link the participants in the pair-production and bremsstrahlung processes would appear to suggest that the addition of new individuals to the cascade in an uncorrelated manner is inadequate. We have included the Poisson law in our discussion because of its broad applicability and its unique and fundamental property

$$\langle k \rangle = V(k) = M_3(k) . \quad (2)$$

Here V and M_3 denote the variance and third central moment of the distribution, respectively. We employ the Fano factor, defined as the variance-to-mean ratio, as an elementary measure of event dispersion.[8] The Fano factor $F_k(T)$ of the above Poisson counting distribution provides a convenient standard for comparison with other distributions, as it is simply given by

$$F_k(T) = V(k)/\langle k \rangle = 1 . \quad (3)$$

The inclusion of the primary particle in the shower statistics can be easily effected by a unit translation of the counting distribution of the secondaries

$$P(m;T) = P(k = m - 1;T) , \quad (4)$$

yielding, in the above case, a shifted-Poisson (one-parameter) distribution for the shower size or multiplication m. Thus,

$$P(m;T) = \langle k \rangle^{m-1} \exp(-\langle k \rangle)/(m-1)! ,$$
$$m \geq 1 , \quad (5)$$

where the mean value of the shifted distribution, $\langle m \rangle = \langle k \rangle + 1$, reflects the deterministic addition of the primary particle to the fluctuating number of secondary events. Since the count variance is unchanged, the Fano factor becomes

$$F_m(T) = (\langle m \rangle - 1)/\langle m \rangle = 1 - \exp(-a) ,$$
$$a \geq 0 , \quad (6)$$

where we have set $\langle m \rangle = \exp(a)$. The reason for this choice of notation will soon be apparent. Note that the Fano factor now approaches unity asymptotically with increasing a as the primary's determinism becomes largely obscured by the growing fluctuations of secondary counts.

The assumptions of this simple model were challenged by Furry[9] soon after their publication. In his landmark paper, Furry considered a pure-birth model whereby each cascade particle may directly breed daughters via a fixed birth rate λ. The multiplication then follows a shifted-geometric (shifted-Bose-Einstein) one-parameter counting distribution

$$P(m;T) = \exp(-\beta)[1 - \exp(-\beta)]^{m-1} ,$$
$$\beta \geq 0, \ m \geq 1 . \quad (7)$$

The statistics up to third order are

$$\langle m \rangle = \exp(\beta) , \quad (8a)$$

$$V(m) = \langle m \rangle^2 - \langle m \rangle , \quad (8b)$$

$$M_3(m) = \langle m \rangle - 3\langle m \rangle^2 + 2\langle m \rangle^3 ; \quad (8c)$$

the Fano factor is, therefore,

$$F_m(T) = \exp(\beta) - 1 . \quad (9)$$

Here the unitless branching parameter $\beta = \lambda t_s$ is the product of the growth rate λ and the clusters playout time t_s. Hence when $t_s = 0$ there is no branching, since only the primary event is present, but the mean multiplication subsequently grows exponentially with t_s. The quantity a in Eq. (6) can now be viewed as an equivalent branching-process parametrization of the shifted Poisson distribution.

The above counting distribution was independently derived earlier by Yule in connection with his elegant mathematical treatment of evolution.[10] An exact parallel follows if we identify the shower population as the number of species within a genus and if we let the branching mechanism represent random mutations. The shifted-Bose-Einstein counting distribution is, therefore, also known as the Yule-Furry distribution, and the point process from which it is derived is termed a Yule-Furry process.

Finally, we examine the use of a related model whose mathematical properties resemble those of the Yule-Furry distribution. This is the one-parameter logarithmic counting distribution

Statistical properties of cosmic-ray showers at ground level determined from photomultiplier-tube background registrations

Malvin C. Teich

Center for Telecommunications Research and Department of Electrical Engineering.
Columbia University, New York, New York 10027

Richard A. Campos

Columbia Radiation Laboratory and Department of Applied Physics, Columbia University, New York, New York 10027

Bahaa E. A. Saleh

Department of Electrical and Computer Engineering, University of Wisconsin, Madison, Wisconsin 53706
(Received 10 March 1987)

The fluctuations of cosmic-ray particles resulting from extensive air showers at ground level are well described by the two-parameter Poisson-driven Yule-Furry and negative-binomial counting probability distributions. The background signal from a single photomultiplier tube has been used to experimentally verify these results with remarkable precision, in spite of the simplicity of the underlying pure-birth stochastic process. Counting distributions from three different photomultiplier-tube detectors operated in the dark are presented, together with the theoretical predictions. Probability distributions of interevent times have also been obtained and these are found to be consistent with the observed clustering properties at the detector output. Our approach is expected to be of importance in quantum optics where cosmic-ray-shower particles can pose a significant limitation on the detection of squeezed light.

I. INTRODUCTION

The rich historical legacy embodied by cosmic-ray research began in 1912 when the extraterrestrial origin of ionizing radiation in the atmosphere was inferred from the pioneering balloon experiments of Victor Hess. By recording their tracks in triggered cloud chambers, Rossi[1] and others demonstrated that some cosmic rays at ground level can initiate cascades and are stopped in thin lead plates, while others continue to penetrate through much thicker layers of the same material without multiplicative effects; hence the designations soft and hard components respectively. The penetrating power of the latter accounted for the first time the existence of carriers of ultrahigh energies and heralded the search for new elementary particles, thus opening an important chapter in modern physics.

The fate of primary cosmic-ray particles from space determines their dual collective behavior in lead, as is now well understood. The primary particles, consisting mostly of protons, but also after entering Earth's atmosphere and fragment into metastable species such as π mesons which, governed by the Yukawa strong force beyond the interaction range, if not captured, the charged pions decay into muons that collide infrequently with the ground as the penetrating face of the hard radiation. Muons may also disintegrate over a longer time span into electrons or positrons with their initial charge state. The neutrinos, on the other hand, are much more unstable than their charged counterparts

and decay swiftly into photon pairs.

These resulting photons and the less abundant muon-derived electrons are seeds of the atmospheric avalanche effect, and, thus, collectively encompass the multiplicative nature of cosmic rays. A single high-energy photon in matter may materialize into an electron-positron pair, whose members in turn radiate further photons through bremsstrahlung collisions. A cyclic mechanism is thus available to initiate and sustain an electromagnetic cascade where the energy of the source particle is partitioned among its offspring until the individual share is low enough for Compton scattering, ionization losses, and other processes to dominate and dilute the avalanche. If the source is sufficiently energetic, however, the family survives down to ground level where it can be detected as an extensive air shower.[3]

It is the statistical behavior of this cascading component of the cosmic-ray progeny that is of principal interest in this paper. We focus on the fluctuations in shower size that arise from the inherent randomness of the multiplication processes, rather than on the distribution of secondary-particle energies. The mathematical description is rooted in stochastic branching-process theory.

From the unfolding history of a primary cosmic ray in the atmosphere, we can clearly identify at least three levels of randomness. They represent the fluctuations in primary arrivals, in the number of cascade-initiating sources per primary, and in the population of each such cascade. For greatest simplicity, however, we eliminate the intermediate level of random behavior and assume

END

DATE
FILMED

4-88
DTIC